

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 20, 2005, 18:16:40 ; Search time 118 Seconds

(without alignments)  
1358.312 Million cell updates/sec

Title: US-10-017-471b-17

Perfect score: 1645  
Sequence: 1 MPEAVVLINSASDANSIEQT.....GHODSLVPLTTLGPAFSG 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_03.\*

1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1634.5	99.4	314	2	086851
2	1023.5	62.2	301	2	AFSA STRGR
3	602	36.6	309	2	056152
4	562	34.2	294	2	024738
5	524.5	31.9	331	2	083X19
6	510	31.0	317	2	084H95
7	480.5	29.2	291	2	024740
8	319.5	19.4	345	2	082KU4
9	271.5	16.5	353	2	08JN90
10	132	8.0	1199	2	07NCE7
11	120.5	7.3	597	2	0727P2
12	118.5	7.2	563	2	08N510
13	118.5	7.2	648	1	Y084 HUMAN
14	109.5	6.7	1762	2	030480
15	106	6.4	372	2	06Z2G9
16	105.5	6.4	942	2	09JL14
17	104.5	6.4	607	2	082M25
18	103.5	6.3	671	2	0925X4
19	103.5	6.3	1089	2	07X2F4
20	101.5	6.2	1002	1	NM3B RAT
21	101	6.1	247	2	06NAK9
22	101	6.1	500	2	089LE3
23	101	6.1	548	2	064QC7
24	101	6.1	6396	2	09KID7
25	100.5	6.1	292	2	082HN2
26	100.5	6.1	777	2	0826D8
27	100	6.1	1292	2	0758Q9
28	99.5	6.0	802	2	08KJH4
29	99	6.0	425	2	084WM4
30	99	6.0	549	2	082B80
31	99	6.0	4391	1	PGBM_HUMAN

32	98.5	6.0	193	2	069PT0	069PT0
33	98.5	6.0	2747	2	09L800	09L800
34	98	6.0	259	2	075L44	075L44
35	98	6.0	1408	2	06X125	06X125
36	97.5	5.9	274	2	062D39	062D39
37	97.5	5.9	274	2	063M77	063M77
38	97.5	5.9	4557	2	08CJN6	08CJN6
39	97	5.9	1480	2	08GC13	08GC13
40	96.5	5.9	263	2	09AD74	09AD74
41	96.5	5.9	495	1	A1BG HUMAN	A1BG HUMAN
42	96.5	5.9	5060	2	052545	052545
43	96	5.8	504	2	082DTS	082DTS
44	96	5.8	725	2	07SEH9	07SEH9
45	96	5.8	1802	2	06VIN7	06VIN7

## ALIGNMENTS

RESULT 1	ID	086851	PRELIMINARY;	PRT;	314 AA.
AC	086851				
DT	01-NOV-1998	(TREMBLrel. 08, Created)			
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)			
DT	25-OCT-2004	(TREMBLrel. 28, Last annotation update)			
DE	ScdA protein.				
GN	Name=ScdA; ORFNames=SCAH10.31;				
OS	Streptomyces coelicolor.				
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OC	Streptomycinae; Streptomycetaceae; Streptomycetes.				
OX	NCBI_TaxID=1902;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Subacteria, and M145;				
RA	Takano E., Chakraborty R., Nihira T., Yamada Y., Bibb M.;				
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=A3(2) / M145;				
RA	MEHLIN=21996410; PubMed=12000953; DOI=10.1038/417141a;				
RA	Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,				
RA	Thomson N.R., James K.D., Harris D.B., Quail M.A., Kleser H.,				
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,				
RA	Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,				
RA	Huang C.-H., Kleser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,				
RA	Rabinowitz E., Rajadream M.A., Rutherford K.M., Rutter S.,				
RA	Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,				
RA	Warren T., Wietzorek A., Woodward J.R., Barrell B.G., Parkhill J.,				
RT	Hopwood D.A.;				
RT	"Complete genome sequence of the model actinomycete Streptomyces				
RT	coelicolor A3(2)";				
RL	Nature 417:141-147(2002).				
DR	EMBL; AJ007731; CAA07627.1;				
DR	EMBL; AL939127; CAB60185.1;				
DR	InterPro; IPR005509; AFSA_repeat.				
DR	Pfam; PF03756; AFSA; 2.				
KM	Complete proteome.				
SO	SEQUENCE 314 AA; 33747 MW; 0F74C9669B6CF389 CRC64;				
Query Match	99.4%; Score 1634.5; DB 2; Length 314;				
Best Local Similarity	99.7%; Pred. No. 2.7e-131;				
Matches	313; Conservative 0; Mismatches 0; Indels 1; Gaps 1;				
OY	1	MPEAVVLINSASDANSIEQTALPPVPMALVHRTVQDAFPVSWIRKGGDRSVTVLPHDH	60		
DB	1	MPEAVVLINSASDANSIEQTALPPVPMALVHRTVQDAFPVSWIRKGGDRSVTVLPHDH	60		
OY	61	PEFAPVHGDRDPLLIETTRQAAMLVFHHGYGVPGVTHFLM-TLDYTCCHLDHGVSGEV	119		
DB	61	PEFAPVHGDRDPLLIETTRQAAMLVFHHGYGVPGVTHFLM-TLDYTCCHLDHGVSGEV	120		
OY	120	ALEVEVACSQLKRRGGQPVQGVDMVRRAGRLAATGATATRTFTSPQVYRRMGDFATP	179		

```

Db 121 ALEVEVACSOQKTFGGQGVQGVMAVRAARLATAATGATRTFSPOVYRRMRGDPATNP 180
Qy 180 TASVGTGAPVPAARAGRTDEDEVLSASSQODTWRLRVDTSHPTLFORPNHVPQMLLLE 239
Db 181 TASVGTGAPVPAARAGRTDEDEVLSASSQODTWRLRVDTSHPTLFORPNHVPQMLLLE 240
Qy 240 AARQAACLVTPGAPFVPVPSIGTRFVRVAFEDSPCMIQATVRPGPAAGLTTVRVTGHQDGS 299
Db 241 AARQAACLVTPGAPFVPVPSIGTRFVRVAFEDSPCMIQATVRPGPAAGLTTVRVTGHQDGS 300
Qy 300 LVFLTLTSGPAPSG 313
Db 301 LVFLTLTSGPAPSG 314

```

## RESULT 2

```

AFSA_STRGR STANDARD; PRT; 301 AA.
ID AFSA_STRGR
AC P18394;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Possible A-factor biosynthesis enzyme.
GN Name=afsa;
OS Streptomyces griseus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxId=1911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89123125; PubMed=2492509;
RT Horinouchi S., Suzuki H., Nishiyama M., Beppu T.;
RT "Nucleotide sequence and transcriptional analysis of the Streptomyces
RT griseus gene (afsa) responsible for A-factor biosynthesis.";
RT J. Bacteriol. 171:1206-1210(1989).
RN [2]
RN SEQUENCE FROM N.A.
RP Umeyama T.;
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: This is a key enzyme for A-factor (2-iscaptyloyl-3R-
CC hydroxymethyl-gamma-butyrolactone) biosynthesis. A-factor is a
CC diffusible bioregulator that is essential for streptomycin
CC production, streptomycin resistance, and spore formation.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24250; AAA26693.1; -
DR EMBL; AB011413; BA32134.1; -
DR PIR; A32061; A32061.
DR InterPro; IPR005509; Afsa_repeat.
DR Pfam; PF03756; Afsa; 2.
KW PfamId.
SQ SEQUENCE 301 AA; 32706 MW; E93302C7547065E3 CRC64;

```

Query Match 62.2%; Score 1023.5; DB 1; Length 301;  
 Best Local Similarity 66.4%; Pred. No. 3.9e-79;  
 Matches 194; Conservative 26; Mismatches 71; Indels 1; Gaps 1;

```

Qy 23 PVPMAVHRTVQDAFPVSWIKGGRFESVTAVLPDHDFPAPVHGRDRLPLIAETLQ 82
Db 10 PVGIEVHRTREDAPFRPMVRLGRDRFAVEAVLPDPHDFPAPVGGDLADPLLVEAMQ 69
Qy 83 AAMLVHAGYGVGVGHFLMT-LDYTGCHDLGVSGEVALEVEVACSOQKTFGGQGVQ 141
Db 70 AAMLVHAGYGVGVGHFLMT-LDYTGCHDLGVSGEVALEVEVACSOQKTFGGQGVQ 129

```

```

Qy 142 QVDMVARRAGRLATGATRTFSPOVYRRMRGDPATRTASVGTAPVPAARAGRTDED 201
Db 130 RVGMAVHRGDLATGVAATRTFSPOVYRRMRGDPATRTASVGTAPVPAARAGRTDED 189
Qy 202 VLSASSQODTWRLRVDTSHPTLFORPNHVPQMLLLEAARQAACLVTPGAPFVPVPSIGCT 261
Db 190 VLSASSQODTWRLRVDTSHPTLFORPNHVPQMLLLEAARQAACLVTPGAPFVPVPSIGCT 249
Qy 262 RFRVYAEFDSPCMIQATVRPGPAAGLTTVRVTGHQDGSLVFLTLTSGPAPSG 313
Db 250 RFRVYAEFDSPCMIQATVRPGPAAGLTTVRVTGHQDGSLVFLTLTSGPAPSG 301

```

## RESULT 3

```

ID 056152 PRELIMINARY; PRT; 309 AA.
AC 056152;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Jadhvi.
GN Name=jadhvi;
OS Streptomyces venezuelae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxId=54571;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ISP5230;
RX MEDLINE=22787062; PubMed=12904539; DOI=10.1099/mic.0.26209-0;
RA Wang L., Vining L.C.;
RT "Control of growth, secondary metabolism and sporulation in
RT Streptomyces venezuelae ISP5230 by jadhvi(1), a member of the afsa
RT family of gamma-butyrolactone regulatory genes.";
RT Microbiology 149:1991-2004(2003).
DR EMBL; U24659; AAB36582.2; -
DR InterPro; IPR005509; Afsa_repeat.
DR Pfam; PF03756; Afsa; 2.
SQ SEQUENCE 309 AA; 33761 MW; 996935C1D855FF4C CRC64;

```

Query Match 36.6%; Score 602; DB 2; Length 309;  
 Best Local Similarity 45.3%; Pred. No. 4e-43;  
 Matches 130; Conservative 39; Mismatches 110; Indels 8; Gaps 3;

```

Qy 29 VHRTRVQDAFPVSWIKGGRFESVTAVLPDHDFPAPVHGRDRLPLIAETLQ 88
Db 10 VHRADPADIIPTDWTQKQNFVSARVACLSILLSRAGRHDPMLVAETIPEISMLVA 69
Qy 89 HAGYGVPGVGHFLM-TLDYTGCHDLGVSGEVALEVEVACSOQKTFGGQGVQGVMAV 147
Db 70 HAELGVPLDEQFVMMWLSYSDSEALTVDGSLSDVTVVCSDTTRKGRRLRLKTTVVL 129
Qy 148 RRAGRLATGATRTFSPOVYRRMRGDPATRTASVGTAPVPAARAGRTDEDEV 203
Db 130 TRDRLATGSGTARCTISALAYRRKGRMALGRPVPLIGVHP--RLVGRARTDEV 186
Qy 204 LSASSQODTWRLRVDTSHPTLFORPNHVPQMLLLEAARQAACLVTPGAPFVPVPSIGTRF 263
Db 187 LAPGRPDQWQLRVNTAHTTLFRPNHVPQMLLLEAARQAACLVTPGAPFVPVPSIGTRF 246
Qy 264 VRYAEFDSPCMIQATVRPGPAAGLTTVRVTGHQDGSLVFLTLTSGPAPSG 310
Db 247 LRYELDSPCMIKAEVSPTPDPSITTRVTGHQDGSVFRCTLTSPS 293

```

## RESULT 4

```

ID 024738 PRELIMINARY; PRT; 294 AA.
AC 024738;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE BarX.

```

GN Name=barX;  
 OS Streptomyces virginiae.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 NC NCB1\_TaxID=1961;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98037495; PubMed=9371444;  
 RA Kinoshita H., Ipposhi H., Okamoto S., Nakano H., Nihira T., Yamada Y.,  
 RT "Butyrolactone autoregulator in Streptomyces virginiae."  
 RL J. Bacteriol. 179:6986-6993(1997).  
 DR EMBL; AB01608; BAA23611.1;  
 DR InterPro; IPR005509; AfsA\_repeat.  
 DR Pfam; PF03756; AfsA; 2.  
 SO SEQUENCE 294 AA; 32378 MW; 7192024EB68F7892 CRC64;  
 Query Match 34.2%; Score 562; DB 2; Length 294;  
 Best Local Similarity 43.9%; Pred. No. 9.8e-40;  
 Matches 123; Conservative 39; Mismatches 116; Indels 2; Gaps 2;  
 QY 24 VPMALVHRTYQDAFPVSMIRKGGREFVTAVLPHDHFPAVHGDRHDPILLIAETTRQA 83  
 DB 5 VPRELVHRAAAVEVFLTGMSRTAENRFALTQMPRAHSYFTPPVNG-CYDPLASTTKQV 63  
 QY 84 AMLVFAGYGVVGYHFLM-TLDYTCGLDHLGSGEVALEVEVASCQLKRGQGPVQGO 142  
 DB 64 GTLSHAEFGVSPGQFLMDLHNSVREQAGVGAPADLELDVICSIRRRGRRLAGMR 123  
 QY 143 VDMAVRAAGRLAATGTATRTFTSPQVYRMRGDFATPTASVGTAPVBARAGRTREDEV 202  
 DB 124 YEVLTCGGQVYATGGAFFDCTSPAVYQRLRGDRGATGVRLPQPLAPASVGRFLTTDV 183  
 QY 203 VLSSASQODTRRLRPVDTSHPTLFPQPNHVGEMLLLEAARQAACLVTPAPFVBSIGTR 262  
 DB 184 VLSTERLEWQLRDEQHPVLFDPVHVPMVLMESARQAQAIDSPRFLPTWSE 243  
 QY 263 FVRVAFDPSPCMIQATVPRGPAAGLTTRVATGHODGSLVF 302  
 DB 244 FSRVAFDPSPCMIQATVPRGPAAGLTTRVATGHODGSLVF 283  
 RESULT 5  
 Q83X19 PRELIMINARY; PRT; 331 AA.  
 AC Q83X19;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE A-factor biosynthesis protein AfsA homolog.  
 OS Streptomyces rochei (Streptomyces parvulus).  
 OC plasmid pSLA2-L.  
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 CX NCB1\_TaxID=1928;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=7434AN4;  
 RX MEDLINE=22676866; PubMed=12791134;  
 RA Mochizuki S., Hiratsu K., Suwa M., Ishii T., Sugino F., Yamada K.,  
 RA Kinashi H.;  
 RT "The large linear plasmid pSLA2-L of Streptomyces rochei has an  
 RT unusually condensed gene organization for secondary metabolism."  
 RL Mol. Microbiol. 48:1501-1510(2003).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=7434AN4;  
 RX MEDLINE=20408175; PubMed=10954087;  
 RA Hiratsu K., Mochizuki S., Kinashi H.;  
 RT "Cloning and analysis of the replication origin and the telomeres of  
 RT the large linear plasmid pSLA2-L in Streptomyces rochei."  
 RL Mol. Gen. Genet. 263:1015-1021(2000).  
 RN (3)

RP SEQUENCE FROM N.A.  
 RC STRAIN=7434AN4;  
 RX MEDLINE=20231737; PubMed=10767533; DOI=10.1016/S0378-1119(00)00060-3;  
 RA Suwa M., Sugino H., Sasaoka A., Mori E., Fujii S., Shinkawa H.,  
 RA Niimi O., Kinashi H.;  
 RT "Identification of two polyketide synthase gene clusters on the linear  
 RT plasmid pSLA2-L in Streptomyces rochei."  
 RL Gene 246:123-131(2000).  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=7434AN4;  
 RX MEDLINE=98031144; PubMed=9836424;  
 RA Kinashi H., Fujii S., Hatani A., Kurokawa T., Shinkawa H.;  
 RT "Physical mapping of the linear plasmid pSLA2-L and localization of  
 RT the eryA and actI homologs."  
 RL Biosci. Biotechnol. Biochem. 62:1892-1897(1998).  
 DR EMBL; AB088224; BAC76543.1;  
 DR InterPro; IPR005509; AfsA\_repeat.  
 DR InterPro; IPR002086; Aldhyd dehydrg.  
 DR InterPro; IPR000623; Shik\_kinase.  
 DR Pfam; PF03756; AfsA; 2.  
 DR PROSITE; PS00687; ALDEHYDE\_DEHYDR\_GLU; UNKNOWN\_1.  
 DR PROSITE; PS01128; SHIKIMATE\_KINASE; UNKNOWN\_1.  
 SO SEQUENCE 331 AA; 36788 MW; 1E35F5BECFA22068 CRC64;  
 Query Match 31.9%; Score 524.5; DB 2; Length 331;  
 Best Local Similarity 38.4%; Pred. No. 1.8e-36;  
 Matches 118; Conservative 44; Mismatches 136; Indels 9; Gaps 5;  
 QY 10 SASDANSIEQTALEFPMALVHRTVQDAFPVSMIRKGGREFVTAVLPHDHFPAVHGCD 69  
 DB 23 SPQDTSWLTPLTTVPREYVHRAIAEVFLTRCTRIHETRLTLGCPRAATFTLSPGR 82  
 QY 70 RHDPLLAETRLQAAMLVFAGYGVVGYHFLM-TLYTTCGLDHLGSGEVALEVEVAC 128  
 DB 83 RHDPMQIAETRRQVGLHIAHAEFDPVPLGHHFIMWDMFVSVEHLGVGRPTDLDVEATC 142  
 QY 129 SQLKFRGGQPVQGVDMANVRAGRLAATGTATRTFT--SPQVYRMRGDFATPTASVGT 186  
 DB 143 VDVVRRRGKLVFELVLTTERDGHVANGG--RFTCTTEAMYRLRSAPATTAAQAS 200  
 QY 187 ---APVPAARAGRTREDEVVLSSASQODTRRLRPVDTSHPTLFPQPNHVPQMLLEAARQ 243  
 DB 201 HQPAPLEPSDGRKARPDDVILAPGAPNRRLNMDTSHPLFDHSGHVPQMLLEAARQ 260  
 QY 244 AACLVTP-APVDSIGTRFVRYAEPDSCWQATVPRGPAAGLTTRVATGHODGSLVF 302  
 DB 261 AACALLPPGSTLLIPATVSTERRRYVEFTSPCMIASGLAVTSGTFALITGRDDDEVF 320  
 QY 303 LTTLSGP 309  
 DB 321 TARISGP 327  
 RESULT 6  
 Q84H95 PRELIMINARY; PRT; 317 AA.  
 AC Q84H95;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Transcriptional regulator.  
 OS Streptomyces carzinostaticus subsp. neocarzinostaticus.  
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 CX NCB1\_TaxID=167636;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15944;  
 RA Liu W., Nonaka K., Nie L., Bae J., Zazopoulos E., Farnet C.M.,  
 RA Shen B.;  
 RT Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.

CLVT 245



QY 250 GPAPFVPSIGTFRVRAVEFDSPC-----WIGATVRCPPAAGLTT----VRYTGHOD 297  
 Db 262 HEAS--RRTLGDRAFALAGLSLDCAFAGELGETILLSEKSPSSEGVPTBECALRVAAHOG 319  
 QY 298 GSLV 301  
 Db 320 ERL 323

## RESULT 9

Q9JN90 PRELIMINARY; PRT; 353 AA.  
 AC Q9JN90;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)  
 DE Hypothetical protein mmfL (Putative lactone biosynthesis protein, mmfL).  
 GN Name=mmfL; OrderedLocuNames=SCPI.243;  
 OS Streptomyces coelicolor.  
 OG Plasmid pSCPI.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3; PLASMID=PCPI;  
 RX MEDLINE=86112873; PubMed=2828187; DOI=10.1016/0378-1119(87)90378-7;  
 RA Neal R.J., Chater K.F.;  
 RT "Nucleotide sequence analysis reveals similarities between proteins determining methylenomycin A resistance in Streptomyces and tetracycline resistance in eubacteria.";  
 RL Gene 58:229-241(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3; PLASMID=PCPI;  
 RX MEDLINE=85284984; PubMed=2992952;  
 RA Chater K.F., Bruton C.J.;  
 RT "Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered.";  
 RL EMBO J. 4:1893-1897(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3; PLASMID=PCPI;  
 RA Bruton C.J., Wietzorrek A., Hartley N., Woodburn L., Chater K.F.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=2196410; PubMed=12000953; DOI=10.1038/417141a;  
 RA Bentley S.D., Chater K.F., Cerdano-Terraza A.-M., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H., Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang C.-H., Kleiser T., Locke L., Murphy L.D., Oliver K., O'Neill S., Rabbilowitsch E., Rajandream M.A., Rutherford K.M., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J., Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 RN [5]  
 RP EMBL; AJ276673; CAB82865.1; -;  
 DR EMBL; AL590464; CAC36769.1; -;  
 DR InterPro; IPR005509; AfsA\_repeat.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF03756; AfsA; 2.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 KW Complete proteome; Hypothetical protein; Plasmid.  
 SQ SEQUENCE 353 AA; 37658 MW; F7BCC6C6D17B74F CRC64;

Query Match 16.5%; Score 271.5; DB 2; Length 353;

Best Local Similarity 27.9%; Pred. No. 7.8e-15;  
 Matches 95; Conservative 43; Mismatches 123; Indels 79; Gaps 12;

QY 21 ALPVMAIVHRTVRQDAFPVSWIPKGDRESVTVLPHDHPFAPVAGDRHPLLAEVL 80  
 Db 26 ARPVPADVHKAAAEVLLDARPLGENRFAVALMWRNFTLMARATSSPCDPLAAETI 85  
 QY 81 RQAAMLVFHAQYGVVGVHFLMTLDYTCYLDHGVSGEVALEVEVACSQLKFRGQPVQ 140  
 Db 86 ROSAHLSTFCVDVPIGHFVL-----SGLDLDLDLPVWDS-----GRLPVV 127  
 QY 141 GQVDMVARRARLAATGTATT-----RFT--TSQVY--- 169  
 Db 128 -----LDVSTKTTTERRRARAALNADVYAGLRGCALRFEVLARRRYAMI 175  
 QY 170 --RRMGDPATPTVSAGTAPVPAARAGRTREDDVVL--SASSQODT--WRLRVDTSHPTLF 225  
 Db 176 RDRARRARRPAQQAAGAAATAPLPETVGFHDDLVLATYAGLPDTYMQRLRRDHYLF 235  
 QY 226 QRPNDVPGMLLEAPAOAACLVTGPAP--FVP-----SIGTRFVRAVEFDSPCWIOAT 278  
 Db 236 DHESDHISGMALIEACQAAATATLPAPGAFGRQVALTVASGYAFGLDSP--VTIT 293  
 QY 279 VRP-----GPAAGLTVRVTVGHQDGLVFLTISGPAFSG 313  
 Db 294 TLPAAHGHSPPSGTTRTLQTPAQSSRTLITATVTTTTAG 333

## RESULT 10

Q7NCZ7 PRELIMINARY; PRT; 1199 AA.  
 AC Q7NCZ7;  
 DT 01-MAR-2004 (TRENBLrel. 26, Created)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE G112829 protein.  
 GN OrderedLocuNames=g112829;  
 OS Gloeobacter violaceus.  
 OC Bacteria; Cyanobacteria; Chroococcales; Gloeobacter.  
 OX NCBI\_TaxID=33072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PCC 7421;  
 RX MEDLINE=22977040; PubMed=14621292;  
 RA Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyaehita H., Tsuchiya T., Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimpo S., Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of Gloeobacter violaceus PCC 7421, a cyanobacterium that lacks thylakoids.";  
 RL DNA Res. 10:137-145(2003).  
 RN [2]  
 RP EMBL; AP006578; BAC90770.1; -;  
 DR GO; GO:0048037; F:cofactor binding; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR009081; ACP\_like.  
 DR InterPro; IPR001227; Ac\_transferase.  
 DR InterPro; IPR005509; AfsA\_repeat.  
 DR InterPro; IPR00794; Phepanteth\_synth.  
 DR InterPro; IPR006163; Phepanteth\_bind.  
 DR InterPro; IPR006162; Pnante S.  
 DR Pfam; PF00698; Acyl\_transfer\_1; 1.  
 DR Pfam; PF03756; AfsA; 1.  
 DR Pfam; PF00109; ketoacyl-synt; 1.  
 DR Pfam; PF02801; ketoacyl-synt C; 1.  
 DR Pfam; PF00550; PP-binding; 1.  
 DR PROSITE; PS50075; ACP\_DOMAIN; 1.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.  
 KW Complete proteome; Transferase.  
 SQ SEQUENCE 1199 AA; 130726 MW; 1554AAE63CA7D543 CRC64;



Db 110 ELDCCSSLDHPTDQKLIPEFIKIOEASOGKIFGVLP---QYHSSVNSAGSSAPVSTA 166.  
 QY 160 TTRFTSPQVYRRMGDPFATPTASVGTAPVPAARAGTRDE-----DVL-----SA 206  
 Db 167 NSTEDARARA-KNARDHSLNENKPGTGDVCSAPAGRNQSPSSGPRGEPFLAKQPSSP 225  
 QY 207 SSQODTWLRVD-----TSHP-----TLFORPNDH 231  
 Db 226 SGEGBGGLSPQVSKTLDGPESNPLEVHEBPLSGKMEIFTLFNPKSXH 274

## RESULT 13

Y084\_HUMAN STANDARD; PRT; 648 AA.  
 AC 014659;  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 05-JUN-2004 (Rel. 44, Last annotation update)  
 DT Hypothetical protein KIAA0084 (HA2022) (Fragment).  
 GN Name=KIAA0084;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=95308325; PubMed=7788527;  
 RA Nagase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,  
 RA Tabata S., Ishikawa K.-I., Kawarabayashi Y., Kotani H., Nomura N.;  
 RT "Prediction of the coding sequences of unidentified human genes. III.  
 RT The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by  
 RT analysis of cDNA clones from human cell line KG-1.";  
 RL DNA Res. 2:37-43(1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D42043; BAA07644.1; -  
 KM Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 648 AA; 70463 MW; 88F68A62D6C1CDA1 CRC64;

Query Match 7.2%; Score 118.5; DB 1; Length 648;  
 Best Local Similarity 27.1%; Pred. No. 0.19;  
 Matches 62; Conservative 22; Mismatches 82; Indels 63; Gaps 11;

QY 50 FSVTAIVLPHDHPFAPVAGDRHDP-----LTAETLRQAAMLVFNAGYGVPGYHFM 102  
 Db 147 FSLAL-----HPVQPTHERREKTPLEHIFRAILIKKTORSQKTDLHNGY-----IL 194  
 QY 103 TLDTTCHLDHGVSEVAVEL--EVEVAVSQ-LKRRGGQVPGQVDMAVRARGLAATTA 159  
 Db 195 ELDCSSLDHPTDQKLIPEFIKIOEASOGKIFGVLP---QYHSSVNSAGSSAPVSTA 251  
 QY 160 TTRFTSPQVYRRMGDPFATPTASVGTAPVPAARAGTRDE-----DVL-----SA 206  
 Db 252 NSTEDARARA-KNARDHSLNENKPGTGDVCSAPAGRNQSPSSGPRGEPFLAKQPSSP 310  
 QY 207 SSQODTWLRVD-----TSHP-----TLFORPNDH 231  
 Db 311 SGEGBGGLSPQVSKTLDGPESNPLEVHEBPLSGKMEIFTLFNPKSXH 359

RESULT 14  
 030480 PRELIMINARY; PRT; 1762 AA.

AC 030480;  
 DT 01-JAN-1998 (TREMELREL. 05, Created)  
 DT 01-JAN-1998 (TREMELREL. 05, Last sequence update)  
 DT 01-MAR-2004 (TREMELREL. 26, Last annotation update)  
 DE PKS module 2.  
 OS Streptomyces hygroscopicus.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 OC NCBI\_TaxID=1912;  
 RN (1)

RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29253;  
 RX MEDLINE=98085969; PubMed=9426000; DOI=10.1016/S0378-1119(97)00450-2;  
 RA Ruan X., Staasi D., Iax S.A., Katz L.;  
 RT "A second type-I PKS gene cluster isolated from Streptomyces  
 RT hygroscopicus ATCC 29253, a rapamycin-producing strain.";  
 RL Gene 203:1-9(1997)  
 DR EMBL; AF007101; AAC38062.1; -  
 DR PIR; T03222; T03222.  
 DR GO; GO:0048037; F:cofactor binding; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR009081; ACP-like.  
 DR InterPro; IPR001227; Ac transferase.  
 DR InterPro; IPR002198; ADH short.  
 DR InterPro; IPR000794; Ketocacyl synth.  
 DR InterPro; IPR006163; Phosphatethyl bind.  
 DR Pfam; PF00698; ACP\_trans\_1; 1.  
 DR Pfam; PF00106; adh\_short; 1.  
 DR Pfam; PF00109; ketocacyl-synt; 1.  
 DR Pfam; PF02801; ketocacyl-synt\_C; 1.  
 DR Pfam; PF00550; PP-binding; 1.  
 DR PROSITE; PS50075; ACP\_DOMAIN; 1.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 KM Transferase.  
 SQ SEQUENCE 1762 AA; 189185 MW; 15BDE12851743835 CRC64;

Query Match 6.7%; Score 109.5; DB 2; Length 1762;  
 Best Local Similarity 24.6%; Pred. No. 3.6;  
 Matches 93; Conservative 33; Mismatches 162; Indels 91; Gaps 19;

QY 10 SASDANSIEOTLAIVPMALVHRTVODAFVSMIPKGGDRFSVAVLPHDHPFAPV--- 66  
 Db 697 SIALNPGHATVLSGDRITLHIAITQANTKTNML-NVSHAFFSLMQPILQPTTTLATL 755  
 QY 67 HGRHNDPLIAETLRQAAMLVFNAGYGVPGYHFMTLDTYCHLDHGVSEVAVELV- 124  
 Db 756 THRPHTPLISMALATGTHPDTH-----WTQHITAPVAYTDTLHLLHHNGITVLEIG 809  
 QY 125 EVAVCSQKRRGGQVPGQVDMAVR-----AGRLAATG-----TATTRFT 164  
 Db 810 PDTTLTALA-RTTLPTTHLIPTRRNHNEVRSTENALGVFSGSHVDWRALPTQRRT 868  
 QY 165 SPQVYRRMGDPFATPTASVGTAPVPAARAGTRDEDEVVLSASQODTWLR---RVDT 220  
 Db 869 SLPTTPQRDRDFWLHDA-GGALEVBGAGLG-TTDHPLLAIVTVADTGLIISGRISTS 926  
 QY 221 HPTLFORPNDH-----VPGMLLEAARQA-----CLVTGPAPFVP 256  
 Db 927 THPPL-----TDHTVNGVIVIPGATLIDLAALAAEHDTHTVDELVINTPLATHTPSQS 982  
 QY 257 SIG-----GRFVRYAEFDP-CYQATV-----RPPAAGLTIVRYTGQDGLVLFT 305  
 Db 983 TVGAEITDGRPALHSDRDTGTWTRHTTGLSNQTHPAELSTWSP---DAHQIDLTT 1039  
 QY 306 L-----SGPFGS 313  
 Db 1040 AYQQLADTGLHYGPAPFG 1057

RESULT 15

```

062ZG9
ID 062ZG9 PRELIMINARY; PRT; 372 AA.
AC 062ZG9;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Prephenate dehydrogenase.
OS Actinoplanes teichomyceticus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micromonosporineae; Micromonosporaceae; Actinoplanes.
OX NCBI_TaxID=1867;
RN [1]
RP SEQUENCE FROM N.A.
RA Li T.L., Huang F., Haydock S.F., Mironenko T., Leadlay P.F.,
RA Spencer J.B.;
RT "Biosynthetic gene cluster of the glycopeptide antibiotic telcoplanin:
RT Characterization of two glycosyltransferases and the key
RT acyltransferase."
RL Chem. Biol. 11:107-119(2004).
RL EMBL; AJ632270; CAG15036.1; -.
DR GO; GO:0047794; F:cyclohexadienyl dehydrogenase activity; IEA.
DR GO; GO:0004665; F:prephenate dehydrogenase (NADP+) activity; IEA.
DR GO; GO:0008977; F:prephenate dehydrogenase activity; IEA.
DR GO; GO:0005571; F:tyrosine biosynthesis; IEA.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR008235; PDH_fbk_insens.
DR InterPro; IPR003099; Prephen_dehydrog.
DR Pfam; PF02153; PDH; 1.
DR PIRSF; PIRSF006786; PDH_fbk_insens; 1.
SQ SEQUENCE 372 AA; 37482 MW; 7DE272A05EDA8D79 CRC64;

Query Match 6.4%; Score 106; DB 2; Length 372;
Best Local Similarity 26.4%; Pred.No.1.1;
Matches 57; Conservative 22; Mismatches 71; Indels 66; Gaps 10;

QY 109 HUDHGVSEVVALEVEVACSQLKFRGQPVQGVDMVAVRAGRLAATG-----T 158
DB 83 YTDVAGVKGQPERDVVRLASDPASVYVGHHPMAGR-----ERSGPLAATADLPAGKTWVLT 137
QY 159 ATTRFTSPQVYRRMRGDFATPTASVPGTAPVPAARAGRTREDDVVLASSQODTWRLRYD 218
DB 138 PNERGAALER-----ATTLAEICGAPVRL-----DSRAHDTVALTS----- 177
QY 219 TSHPTLFQRPNDHVPQMLLEAARQAACLVGTGPAPFVPSIG-GTRPV-RYAEFDSPCW-- 274
DB 178 -----HVPLM--ASLTAARLHGPAATEALVGGVDRDVTRIAGDPALMTD 222
QY 275 IQATVAPGPAAGLTTVR-----VTGHODG 298
DB 223 IVRSNAPAAVAALRGVRDDLTRLAATDILAGHHPG 258

```

Search completed: July 20, 2005, 22:12:16  
 Job time : 126 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 22:12:24 ; Search time 1184 Seconds  
(without alignment)  
102.621 Million cell updates/sec

Title: US-10-017-471B-17

Perfect score: 1645  
Sequence: 1 MPEAVVILINSASDANSIEQT.....GHQDGLVPLTTTSGPAFSG 313

Scoring table: BLOSUM62  
Gap 10.0 ; Gapext 0.5

Searched: 1736639 seqs, 388188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1645	100.0	313	US-10-017-471A-17	Sequence 17, Appl
2	1634.5	99.4	314	US-10-168-663-14	Sequence 14, Appl
3	1023.5	62.2	301	US-10-168-663-13	Sequence 13, Appl
4	562.5	34.2	314	US-10-168-663-15	Sequence 15, Appl
5	482.5	29.3	291	US-10-168-663-16	Sequence 16, Appl
6	319.5	19.4	345	US-10-156-761-9806	Sequence 9806, Ap
7	288	17.5	57	US-10-017-471A-17	Sequence 14, Appl
8	271.5	16.5	353	US-10-168-663-17	Sequence 17, Appl
9	118.5	7.2	648	US-10-755-889-155	Sequence 455, App
10	106.5	6.5	447	US-10-425-115-285291	Sequence 285291,
11	106.5	6.5	460	US-10-425-114-66345	Sequence 66345, A

12	105.5	6.4	256	16	US-10-425-115-204431	Sequence 204431,
13	104.5	6.2	607	14	US-10-156-761-9374	Sequence 9374, Ap
14	101.5	6.2	987	15	US-09-934-070-2	Sequence 2, Appl1
15	101.5	6.2	987	15	US-10-222-772-2	Sequence 2, Appl1
16	101.5	6.2	987	15	US-10-222-772-60	Sequence 60, Appl1
17	101.5	6.2	1002	10	US-09-934-070-4	Sequence 4, Appl1
18	101.5	6.2	1002	15	US-10-222-772-4	Sequence 4, Appl1
19	101.5	6.2	1002	15	US-10-222-772-58	Sequence 58, Appl1
20	101	6.1	3537	18	US-10-922-282-4	Sequence 4, Appl1
21	101	6.1	6396	10	US-09-940-3168-72	Sequence 72, Appl1
22	100.5	6.1	292	14	US-10-156-761-14784	Sequence 11013, A
23	100.5	6.1	777	14	US-10-156-761-14784	Sequence 14784, A
24	100	6.1	751	16	US-10-425-115-218590	Sequence 218590,
25	99	6.0	315	15	US-10-425-114-38543	Sequence 38543, A
26	99	6.0	365	15	US-10-425-114-38900	Sequence 38900, A
27	99	6.0	430	14	US-10-425-114-72573	Sequence 72573, A
28	99	6.0	549	15	US-10-156-761-13294	Sequence 13294, A
29	99	6.0	559	15	US-10-425-114-54109	Sequence 54109, A
30	99	6.0	700	16	US-10-425-115-218593	Sequence 218593,
31	99	6.0	3588	16	US-10-741-601-378	Sequence 378, App
32	99	6.0	3588	17	US-10-741-601-1104	Sequence 1104, App
33	99	6.0	4346	16	US-10-741-601-377	Sequence 377, App
34	99	6.0	4346	17	US-10-741-601-1103	Sequence 1103, App
35	99	6.0	4347	16	US-10-741-601-376	Sequence 376, App
36	99	6.0	4347	17	US-10-741-601-1102	Sequence 1102, App
37	99	6.0	4370	16	US-10-408-765A-1267	Sequence 1267, App
38	97	5.9	4391	16	US-10-478-451-1	Sequence 1, Appl1
39	97	5.9	4393	15	US-10-231-956A-366	Sequence 366, App
40	97	5.9	4393	17	US-10-741-600-1105	Sequence 1105, App
41	96.5	5.9	495	11	US-09-833-245-1680	Sequence 1680, Ap
42	96.5	5.9	495	17	US-10-473-519-16	Sequence 16, Appl1
43	96.5	5.9	495	18	US-10-416-738-19	Sequence 19, Appl1
44	96	5.8	406	15	US-10-344-738-106	Sequence 106, App
45	96	5.8	504	14	US-10-156-761-12417	Sequence 12417, A

ALIGNMENTS

RESULT 1  
US-10-017-471A-17  
US-10-017-471A-17 Application US/10017471A  
; Publication No. US20030124644A1  
; GENERAL INFORMATION:  
; APPLICANT: Takano, Eriko  
; TITLE OF INVENTION: Antibiotic Production  
; FILE REFERENCE: 0380-P023290S1  
; CURRENT APPLICATION NUMBER: US/10/017,471A  
; PRIOR FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: US 60/242,561  
; PRIOR FILING DATE: 2000-10-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor  
US-10-017-471A-17

Query Match 100.0%; Score 1645; DB 14; Length 313;  
Best Local Similarity 100.0%; Pred. No. 8.4e-154;  
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEAVVILINSASDANSIEQTALPVPMALVHRTRODAFPVSWIPKGDGDRSVTAVALPHDH 60  
DB 1 MPEAVVILINSASDANSIEQTALPVPMALVHRTRODAFPVSWIPKGDGDRSVTAVALPHDH 60  
QY 61 PFPAPVHGDHRDPLLIETTRROAAMLVHAGYGVGVYHFLMTLDYTCCHDLGVSSEVA 120  
DB 61 PFPAPVHGDHRDPLLIETTRROAAMLVHAGYGVGVYHFLMTLDYTCCHDLGVSSEVA 120  
QY 121 ELAEVACSQLKFRGGPVOGVQVMAVRARGLAATGATATRTSPQVYRRMRGDFATPT 180

[illegible]

```

RESULT 2
US-10-168-663-14
; Sequence 14, Application US/10168663
; Publication No. US20040086962A1
; GENERAL INFORMATION:
; APPLICANT: Plant Bioscience Limited
; APPLICANT: Chater, Keith F
; APPLICANT: Bruton, Celia J
; APPLICANT: O'Rourke, Sean J
; APPLICANT: Mietznerrek, Andreas W
; TITLE OF INVENTION: Methods and Materials Relating to Gene Expression
; FILE REFERENCE: 0380-P02909USO
; CURRENT APPLICATION NUMBER: US/10/168,663
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: PCT/GB00/04972
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: GB 9930477.6
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Streptomyces sp.
US-10-168-663-14

```

Query Match	99.4%;	Score 1634.5;	DB 15;	Length 314;
Best Local Similarity	99.7%;	Pred. No. 9.2e-153;		
Matches 313; Conservative	0;	Mismatches 0;	Indels 1;	Gaps 1

Qy	60	MPEAVVLINASAPNSIIEQTALFVPALVHRRVDAPFVSI PKGGRFSEVTALPHDH
Db	60	1 MPEAVVLINASAPNSIIEQTALFVPALVHRRVDAPFVSI PKGGRFSEVTALPHDH
Qy	113	61 PFAPRPNHGDHDELLAETLRQAMLVFAGVGVPVGHFFLM-TLDYCHLDHGSSEV
Db	120	61 PFAPRPNHGDHDELLAETLRQAMLVFAGVGVPVGHFFLMATLDYCHLDHGSSEV
Qy	179	120 AELEVEVACSQLKFRGGOPVQGVQDVMAVRAARLATGATATRTFSPVYRRMGDEATP
Db	180	121 AELEVEVACSQLKFRGGQPVQGVQDVMAVRAARGLATGATATRTFSPVYRRMGDEATP
Qy	233	180 TASYPGTAPVPARAGRTREDEVVLSASSQOOTWRLRVDTSPTLFGPRNDHVEGMILLE
Db	240	181 TASYPGTAPVPARAGRTREDEVVLSASSQOOTWRLRVDTSPTLFGPRNDHVEGMILLE
Qy	295	240 AARQAACTLVGPAPFVPSIGTRFVYVAEFSDFPCWIOATVRGHPAAGLTVRVGHODGS
Db	300	241 AARQAACTLVGPAPFVPSIGTRFVYVAEFSDFPCWIOATVRGHPAAGLTVRVGHODGS
Qy	313	300 LVFTLTLSGPAFSG 313
Db	314	301 LVFTLTLSGPAFSG 314

RESULT 3  
US-10-168-663-13

```

Sequence 13, Application US/10168663
Publication No. US2004006962A1
GENERAL INFORMATION:
APPLICANT: Plant Bioscience Limited
APPLICANT: Chater, Keith F
APPLICANT: Bruton, Celia J
APPLICANT: O'Rourke, Sean J
APPLICANT: Wietzorrek, Andreas W
TITLE OF INVENTION: Methods and Materials Relating to Gene Expression
FILE REFERENCE: 0380-P02909050
CURRENT APPLICATION NUMBER: US/10/168,663
CURRENT FILING DATE: 2002-10-25
PRIOR APPLICATION NUMBER: PCT/GB00/04972
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: GB 9930477.6
PRIOR FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn version 3.0
SEQ ID NO 13
LENGTH: 301
TYPE: PRT
ORGANISM: Streptomyces sp.
US-10-168-663-13

Query Match          62.2%; Score 1023.5; DB 15; Length 301;
Best Local Similarity 66.4%; Pred. No. 2.2e-97;
Matches 194; Conservative 26; Mismatches 71; Indels 1; Gaps 1;

```

Qy	23	PVPMALVHSTRVQDAFPPVSMLEPKGDRGSVYAVLPHDHPFAPAHGRHOBLEIAETLRQ	82
Db	10	PVGIEVHHTREDPAFPRNMVWLRGRDRAVEAVLPHDHPFAPVGGDDLDHDLVAEMRQ	69
Qy	83	AAMLVYHAGYCPVGYHFLMT-LDYTCGLDHLGVSGVALEVAEVAASQLKFRGGQVQG	141
Db	70	AAMLAFAHAGYGLGYHFLMTLEDVDCRHHGVGGSEPTLGLGVFCSDLKMRAGIPAQG	129
Qy	142	QVDMVAARRAGRIAACTGATATTPSTSPQVYRRMRGDPATPASPSTAPVAPAAAGRTDED	201
Db	130	RVMVAHHRGDRILAAISVAAITRSTSKAYRRMRGDPVPEGISLEPTAVPASPAGRAVED	189
Qy	202	VVLSSAQDDTWLRVDVTSHPILLFORPNDHVPKMLLEAAFOAACLVTPRAPVPSIGGT	261
Db	190	VVLSTGRGSEVETLRVDTHPTLFORPNDHVPKMLLEAAQACLVAPGAGIVPEART	249
Qy	262	RFRVAYLEPSPCMIAQTVTRPGPAALTVYRVYGHODSLVLETLTSGPASPQ	313
Db	250	RFRVAYLEPSPCMIAQTVTRPGPAEDTVYRVYGHODGTVSVLSTSGPRAHQ	301

RESULT 4  
 US-10-168-663-15  
 Sequence 15, Application US/10168663  
 Publication No. US20040086962A1  
 GENERAL INFORMATION:  
 APPLICANT: Plant Bioscience Limited  
 APPLICANT: Chater, Keith F  
 APPLICANT: Bruton, Ceila J  
 APPLICANT: O'Rourke, Sean J  
 APPLICANT: Wietzorrek, Andreas W  
 TITLE OR INVENTION: Methods and Materials Relating to Gene Expression  
 FILE REFERENCE: 0380-P0290US0  
 CURRENT APPLICATION NUMBER: US/10/168,663  
 CURRENT FILING DATE: 2002-10-25  
 PRIOR APPLICATION NUMBER: PCT/GB00/04972  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: GB 9930477.6  
 PRIOR FILING DATE: 1999-12-23  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 15  
 LENGTH: 314  
 TYPE: PRT  
 ORGANISM: Streptomyces sp.

US-10-168-663-15

Query Match 34.2%; Score 562.5; DB 15; Length 314;  
 Best Local Similarity 42.9%; Pred. No. 8.7e-47;  
 Matches 126; Conservative 39; Mismatches 118; Indels 11; Gaps 3;

QY 19 QTALEP-----VPMALVHRTVQDAFPVSWIPKGGDFSVTAVALPHDHPFAPVHGD 69  
 DB 11 QGFPAPASAMTSTVPRELVRRAVAEVFLTGMSTRTAENRFPALTAQMPRAHSYFTPVNG- 69  
 QY 70 RHDPLLAETLRQAMLVFHAAGVVPVGHFLM-TLDTYCHLDLHSGVGAELLEVAVAC 128  
 DB 70 CYDPLAETLRQVGLTSLHAEFVSFGDQPLMDLHHSVRPEQAGVGAAPADLELVIC 129  
 QY 129 SOLKFRGGQPVQGVDAVARRAGLAATGTATFTSPQVYRRKRGCFATPTASVPSTAP 188  
 DB 130 SDIRRRGRRLAGMYETLYGGQVATGGAAPCTSPRAYVQRIRGDRVATGVRPLPQ 189  
 QY 189 VPAARAGRTDEDEVLASQODTWRLRVDTSHPTLFQRPNDHVPKMLLEAARQAACLV 248  
 DB 190 LAPASVGRFLTTVDVLSTATERPLEMQLRVDEQHFLVFDHVPDHPGVMLEMSAQAQAI 249  
 QY 249 TGPAPFPISIGTFFVRYAERDSCWQATVRPQPAAGLTTRVYTGHDGSLVF 302  
 DB 250 DPSRPFPTWRSFSSRYAELDRPCWIOAEPLPADNGDRQVRYGHQDDTTVF 303

## RESULT 5

US-10-168-663-16

; Sequence 16, Application US/10168663  
 ; Publication No. US20040086962A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Plant Bioscience Limited  
 ; APPLICANT: Chater, Keith F  
 ; APPLICANT: Bruton, Celia J  
 ; APPLICANT: O'Rourke, Sean J  
 ; APPLICANT: Metcortek, Andreas W  
 ; TITLE OF INVENTION: Methods and Materials Relating to Gene Expression  
 ; FILE REFERENCE: 0380-202809US0  
 ; CURRENT APPLICATION NUMBER: US/10/168, 663  
 ; PRIOR FILING DATE: 2002-10-25  
 ; PRIOR APPLICATION NUMBER: PCT/GB00/04972  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: GB 9930477.6  
 ; PRIOR FILING DATE: 1999-12-23  
 ; NUMBER OF SEQ ID NOS: 35  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 16  
 ; LENGTH: 291  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces sp.  
 US-10-168-663-16

Query Match 29.3%; Score 482.5; DB 15; Length 291;  
 Best Local Similarity 40.2%; Pred. No. 6.3e-39;  
 Matches 117; Conservative 44; Mismatches 121; Indels 9; Gaps 6;

QY 28 LVHRTVQDAFPVSWIPKGGDFSVTAVALPHDHPFAPVHGDHDPDLIAETLRQAMLV 87  
 DB 1 LVHRTSTQAVLLTQMORLDDARFSTVAPLASHAFETPVGGYDPLMCACETIRQIAYLL 60  
 QY 88 FHAGVGVVGHFLMTDITGHL---DHLGSGVALEEVAVASQKFRGGQPVQGVVD 144  
 DB 61 GHAEPAVFGHQFVL-WDLSSVVRPELRLGVLPAVDLITCEVIRRRGRSLGAGYE 119  
 QY 145 WAVRAGRLAATGATATFTSPQVYRRKRGDFA-TPT-ASVGTAPVPAARAGRTDEVD 202  
 DB 120 AVARRDQGVATGASVTCSPRAYVQRIRPENVLTPRHPRLPLTAPAPQVVALSPDV 179  
 QY 203 VLSASSQODTWRLRVDTSHPTLFQRPNDHVPKMLLEAARQAACLVGPAPFVPSIGTR 262  
 DB 180 VLSPLDENRQWLTVDTHPVLFDHWDHVPGMVLEAARQAASALGRPSFMBLGVAGE 239

QY 263 FVRYAERDSCWQIA--TVRGPAGLTTRVYTGHDGSLVFLTSLGPAF 311  
 DB 240 FKRYVELDACVIESERLFDQVPAE-EVRYVTGHONGELTFVGTAYASY 289

## RESULT 6

US-10-156-761-9806

; Sequence 9806, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156, 761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 9806  
 ; LENGTH: 345  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-9806

Query Match 19.4%; Score 319.5; DB 14; Length 345;  
 Best Local Similarity 32.2%; Pred. No. 1e-22;  
 Matches 98; Conservative 33; Mismatches 140; Indels 33; Gaps 8;

QY 23 PVPMAVHRTVQDAFPVSWIPKGGDFSVTAVALPHDHPFAPVHGDHDPDLIAETLRQ 82  
 DB 28 PVRRELVRSSIAEVFTDVGRTGENAFVGAQMPRDHALYHDPDENGINDPLFAETLRQ 87  
 QY 83 AAMLVFAGYGVVGHFL---MTLDTYCHLDLHSGVGAELLEVAV--ACSQLKFRGGQ 137  
 DB 88 AHFGATYFVGVPVGSFFIQDVSFEIT--DPTALVGAAPLAVLVNGTWTEBRDRGR 144  
 QY 138 PVQGVDAVARRAGLAATGTATFTSPQVYRRKRGDFATPTASVGPAPVPA----- 192  
 DB 145 PAGARLDVLTLDGRCPCGRHTGMLMDRRYRLRG--RPAASGVSPRRAPADKRIA 201  
 QY 193 ---RAGRTDEDEVLASQODTWRLRVDTSHPTLFQRPNDHVPKMLLEAARQAACLV 249  
 DB 202 RPRRVGLRWKDCVLEDRDPQWRRLRVDRDAVLPHPTDHPVPLVMMLSGFQLGHLTV 261  
 QY 250 GPAPFVPSIGTRPVRYAEFDSPC-----WQATVRGPAGLT-----VRVTHOD 297  
 DB 262 HEAS--RRTIGDBAFALAGSLDCAAFGEIGETILSLKRGPSGVPTSCALRVAAHQG 319  
 QY 298 GSLV 301  
 DB 320 ERLI 323

## RESULT 7

US-10-017-471A-14

; Sequence 14, Application US/10017471A  
 ; Publication No. US2003012464A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takano, Eriko  
 ; APPLICANT: Bibb, Mervyn  
 ; TITLE OF INVENTION: Antibiotic Production  
 ; FILE REFERENCE: 0380-P02329US1  
 ; CURRENT APPLICATION NUMBER: US/10/017, 471A  
 ; CURRENT FILING DATE: 2001-10-23  
 ; PRIOR APPLICATION NUMBER: US 60/242,561

PRIOR FILING DATE: 2000-10-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 57  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor  
US-10-017-471a-14

Query Match 17.5%; Score 288; DB 14; Length 57;  
Best Local Similarity 100.0%; Pred. No. 1.1e-20;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEAVVILNSASDANSIECTALPVPMALVHRTRVODAPFVSMTIPKSGRFSVTALP 57  
DB 1 MPEAVVILNSASDANSIECTALPVPMALVHRTRVODAPFVSMTIPKSGRFSVTALP 57

## RESULT 8

US-10-168-663-17  
; Sequence 17, Application US/10168663  
; Publication No. US20040086962A1  
; GENERAL INFORMATION:  
; APPLICANT: Plant Bioscience Limited  
; APPLICANT: Chater, Keith F  
; APPLICANT: Bruton, Ceila J  
; APPLICANT: O'Rourke, Sean J  
; APPLICANT: Wietzorek, Andreas W  
; TITLE OF INVENTION: Methods and Materials Relating to Gene Expression  
; FILE REFERENCE: 0380-P02909US0  
; CURRENT APPLICATION NUMBER: US/10/168,663  
; CURRENT FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: PCT/GB00/04972  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: GB 9930477.6  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: Patent In version 3.0  
; SEQ ID NO 17  
; LENGTH: 353  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor  
US-10-168-663-17

Query Match 16.5%; Score 271.5; DB 15; Length 353;  
Best Local Similarity 27.9%; Pred. No. 6e-18;  
Matches 95; Conservative 43; Mismatches 123; Indels 79; Gaps 12;

QY 21 ALPVMALVHRTRVODAPFVSMTIPKSGRFSVTALPHDHFFFAVHGDRHPLIAETL 80  
DB 26 APLEPPADVHKAAAEVLLTDPARPGENRFAVAALMPRTFLAHRATSSPCDPLAAETI 85  
QY 81 ROAAMLVFAGVGVYFLMTLDYTCYLDHLSGSEAELEVEVACSQLKFRGGOVPQ 140  
DB 86 ROSAHLHSTFCDVPIGHFVL-----SGLDLDLDLVWMS-----GPLFV 127  
QY 141 GQDVAVRRAGRLAATGATTT-----RR--TSPQVY-- 169  
DB 128 -----LDVTSTTTNNPRMARALNADYVAGLHGRCAIREVLAPRRYMI 175  
QY 170 ---RRNRGDPATPTASVGTAPVAPABAGTRDEYVL--SASSQOQT--WRLYDTSHPTLF 225  
DB 176 RDRARRAERPPAOQAAGATATLPPETVGFHDHLHLAQAQGLPTAMQRLRRDHPVLF 235  
QY 226 QRPNDHVPMILLLEAAROAACLVGTGPAP--FVP-----SIGTRFVRVAFSPSCMIQAT 278  
DB 236 DHESDHISGALLLEACROATATLTPPAGAFGRVOVALVASSYQAFGEIDSP--VIT 293  
QY 279 VRP-----GPAAGLITVTVTGHDGSLVFLTTISGPAFSG 313  
DB 294 TLPAHGHSPDSGTRTLQTLARQGSRTLTATVTTTTTAG 333

## RESULT 9

US-10-755-889-455  
; Sequence 455, Application US/10755889  
; Publication No. US20040171823A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: PATHWAY  
; FILE REFERENCE: D0284 NP  
; CURRENT APPLICATION NUMBER: US/10/755,889  
; CURRENT FILING DATE: 2004-01-13  
; PRIOR APPLICATION NUMBER: U.S. 60/440,068  
; PRIOR FILING DATE: 2003-01-14  
; PRIOR APPLICATION NUMBER: U.S. 60/469,757  
; PRIOR FILING DATE: 2003-05-12  
; NUMBER OF SEQ ID NOS: 823  
; SOFTWARE: Patent In version 3.2  
; SEQ ID NO 455  
; LENGTH: 648  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-755-889-455

Query Match 7.2%; Score 118.5; DB 16; Length 648;  
Best Local Similarity 27.1%; Pred. No. 0.018;  
Matches 62; Conservative 22; Mismatches 82; Indels 63; Gaps 11;

QY 50 FSVTAVLPHDHPFAPVHGDRHPL-----LIAETLROAAMLVFAGVGVYHFLM 102  
DB 147 FSLAAL-----HPFQPTHEREKTELEHIFRALIKKTDRSQKTDLHNEGYS-----IL 194  
QY 103 TLDTYCHDHLGSGVEAEI--EVENACSQ--LKRGGOPVGOVDMVAPRAGRLLAATGTA 159  
DB 195 ELDCSSLDHPTDQKLIPEFIKKIQEASQGLKFEVGIYF--QYHSSVNSAGSSAPVSTA 251  
QY 160 TTRFTSPQVYRMRGDFATPTASVGTAPVAPABAGTRDE-----DVVL-----SA 206  
DB 252 NSTEDARDA-KNARGDASLENEKPGTGDVCSAAGNQSPEBSSGRGVPLAKQSSP 310  
QY 207 SSQDDTWRLRYD-----TSHP-----TLFQRPNDH 231  
DB 311 SGBDGGELSPQGVSKTLDGPESNPLFVHEEPLSGKMEIFLTFKPKPSH 359

## RESULT 10

US-10-425-115-285291  
; Sequence 285291, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 285291  
; LENGTH: 447  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: WRT4577\_23284C.1.pcp  
US-10-425-115-285291

Query Match 6.5%; Score 106.5; DB 16; Length 447;  
Best Local Similarity 22.6%; Pred. No. 0.17;  
Matches 74; Conservative 23; Mismatches 105; Indels 125; Gaps 14;

QY 62 FPAVHGDRHPLIAETL--ROAAMLVFAGVGVYHFLMTLDYTC-----LDH 112



Db 14 PFLPFGCHISBMLQALHARGLAVTLHTGNAP-----DATRIHELTFFVDIHE 64  
 Qy 113 LGVSGEVALEVEVACSQLK-----FRGGQP-----VQGVDMVAVRAG 151  
 Db 65 SSFPEVTSIGTDIVTQLALNACAPFRREALSLRGGQDVACAVVDDGCVSALRAAH 124  
 Qy 152 RLAA-----TGATTTFTSPQVYRRMGDFATPTASVPGTAPVPAARAGTRD----- 199  
 Db 125 RLGVPAVLVLRDTSAAAT-FSSWLAIPRLRDAGFVPEKEERLDEPVDLERLARADLIRVDG 183  
 Qy 200 -----EDVLSASS-----QODTWRLRVDTSHPT-----LFOR 227  
 Db 184 SDTDALCGFIARVADAVARASGVVINTFERMEASELAKIQRELSRPAFVAGPLHLISQA 243  
 Qy 228 P-----NDHVPGLMLEAARQAACLVTPAPFVPSIGTRFVRYAEFDS 271  
 Db 244 PAEGSLHAPDRGCLAMLDHPRRSVLVYSGSVACVDRG-----AFVEMA----- 288  
 Qy 272 PCWIOA-----TVRPGPAAGLTTV 290  
 Db 289 --WQARSVGSFLMVVRPGLVGVPEV 313

RESULT 11  
 US-10-425-114-66345  
 ; Sequence 66345, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdeng  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(5313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425, 114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 66345  
 ; LENGTH: 460  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LIB4729-040-F10\_P11.pep  
 US-10-425-114-66345

Query Match 6.5%; Score 106.5; DB 15; Length 460;  
 Best Local Similarity 22.6%; Pred. No. 0.18;  
 Matches 74; Conservative 23; Mismatches 105; Indels 125; Gaps 14;  
 Qy 62 FFAPVHGDRHDLIAETL--RQAAMLVFHAGVVPVGHFLMLTDYCH-----LDH 112  
 Db 27 FFLPFGCHISBMLQALHARGLAVTLHTGNAP-----DATRIHELTFFVDIHE 77  
 Qy 113 LGVSGEVALEVEVACSQLK-----FRGGQP-----VQGVDMVAVRAG 151  
 Db 78 SSFPEVTSIGTDIVTQLALNACAPFRREALSLRGGQDVACAVVDDGCVSALRAAH 137  
 Qy 152 RLAA-----TGATTTFTSPQVYRRMGDFATPTASVPGTAPVPAARAGTRD----- 199  
 Db 138 RLGVPAVLVLRDTSAAAT-FSSWLAIPRLRDAGFVPEKEERLDEPVDLERLARADLIRVDG 196  
 Qy 200 -----EDVLSASS-----QODTWRLRVDTSHPT-----LFOR 227  
 Db 197 SDTDALCGFIARVADAVARASGVVINTFERMEASELAKIQRELSRPAFVAGPLHLISQA 256  
 Qy 228 P-----NDHVPGLMLEAARQAACLVTPAPFVPSIGTRFVRYAEFDS 271  
 Db 257 PAEGSLHAPDRGCLAMLDHPRRSVLVYSGSVACVDRG-----AFVEMA----- 301  
 Qy 272 PCWIOA-----TVRPGPAAGLTTV 290

Db 302 --WQARSVGSFLMVVRPGLVGVPEV 326

RESULT 12  
 US-10-425-115-204431  
 ; Sequence 204431, Application US/10425115  
 ; Publication No. US20040214272A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants  
 ; FILE REFERENCE: 38-21(53222)B  
 ; CURRENT APPLICATION NUMBER: US/10/425, 115  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 369326  
 ; SEQ ID NO 204431  
 ; LENGTH: 256  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: MRT4577\_118032C.1.pep  
 US-10-425-115-204431

Query Match 6.4%; Score 105.5; DB 16; Length 256;  
 Best Local Similarity 26.0%; Pred. No. 0.099;  
 Matches 57; Conservative 19; Mismatches 58; Indels 85; Gaps 13;  
 Qy 112 HLGVSGEVALEVEVACS-----QKFRGGQPVQGO--VDMVAVRB--GRLATGATTRFT 164  
 Db 95 HSNVNRGGRVNRPIVCAAVFQLTRELGLKSDGTEIEMLRQAEPISILATSGTT--- 151  
 Qy 165 SPOVYRRMGDFATPTASVPGTAPVPAARAGTRDEDDVLSASSQODTWRLRVDTSHPT 224  
 Db 152 -PAVF-----SCSSAPSTS---AASAGAGAGAVSL-----PLT 181  
 Qy 225 FQRP-NDHVPGLMLEAARQAACLVTPAPFVPSIGTRFVRYAEFDSPCMIOATVRGP 283  
 Db 182 GKRPREEHEP-----AVAVAVAPF-----MATLOARFVA 212  
 Qy 284 AA-GLT-----TVRVTHGDSGLVFLTTLSGA 310  
 Db 213 AAMGLSPAGQAQAQAVASVAAQGH--LNLISVLSGA 248

RESULT 13  
 US-10-156-761-9374  
 ; Sequence 9374, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156, 761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 9374  
 ; LENGTH: 607  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis

US-10-156-761-9374

Query Match 6.4%; Score 104.5; DB 14; Length 607;  
Best Local Similarity 24.0%; Pred. No. 0.41;  
Matches 88; Conservative 28; Mismatches 136; Indels 115; Gaps 17;

QY 3 EAVVLINSASDANSIE-OTALPVFMA-----LVHRTYVD----- 36  
DB 175 EAAVLAAAPDSAPBEPBAALAVPQAGDGLVLRVAVVPASLDYRAHGGVAAALRA 234  
QY 37 -----APVSMIPKGGDRFSVTAVLPDHDPFAPVHGDHR-- 71  
DB 235 FELGAEVIREVTADGLVGRGAAPPT-----GKKMQATISRP--DRPHALVCANADESEP 287  
QY 72 -----DPLIAETLRQAAMLVFHAGYGVGVHFLMTLDYTGCHLGLVSGEVA 120  
DB 288 GTFKQVRVMEGDPVALVESMTIAAAYV-----GAHQGLYLKRG--EYPRALSRM----- 334  
QY 121 ELEVEVACSQKFRG--GQPVQGG--VDMAVRARGRLAATGTATRTFSQVYRR----- 171  
DB 335 -----SHAEOQARARGLDDVYGGQYARDIEIRGAGAYICGEBETALFNSIEGYRGEPRS 390  
QY 172 -----MGDFATPTA--SVPGTAPV--PAARAGRTREDVVLASSSQDTRRLRYDTSHP 222  
DB 391 KRPPEVEKGLRGRFTAHNNVETLVNVLVLTWGAQAYAIAGTAKSTGKLCVSGSVDRP 450  
QY 223 TLFORPNHDVPGMLLEAA--ROAACTVGTGPAPFPVPSIGTRFVRYAEPDSPCWIAQTV 279  
DB 451 GUYELPFGATLGELELGLAVGERLRAVLGGAA-----GG--FVRPDERDIPLTFEGTR 502  
QY 280 RPPGPAAG 286  
DB 503 QAGTTLG 509

RESULT 14

US-09-934-070-2  
; Sequence 2, Application US/09934070  
; Publication No. US20030092004A1  
; GENERAL INFORMATION:  
; APPLICANT: Lipton, Stuart A.  
; APPLICANT: Zhang, Dongxian  
; APPLICANT: Chatterton, Jon E.  
; APPLICANT: Awobuluyi, Marc  
; APPLICANT: Sevarino, Kevin A.  
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS  
; FILE REFERENCE: P-LJ 4900  
; CURRENT APPLICATION NUMBER: US/09/934,070  
; CURRENT FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 987  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-09-934-070-2

Query Match 6.2%; Score 101.5; DB 10; Length 987;  
Best Local Similarity 23.8%; Pred. No. 1.6;  
Matches 70; Conservative 31; Mismatches 96; Indels 97; Gaps 13;

QY 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTVQDAFPVSWIPKG-----GDRFSVTA 54  
DB 217 VPAAVVLGCGSTARAEVLEAAPPQOWILGTPLPAEALPTTGTPGVALAGETEQHSLEA 276  
QY 55 VLPHDHPFAPVHGRHDP--LIAETLRQAAMLVFHAGYGVGVHFLMTLDYTGCHLDHL 113  
DB 277 VV-----HDMVELVAQALSSMAL----- 294  
QY 114 GVSGEVALEVEVACSQK-----FRGGQPVQGVDMAVRARGRLAATGTAT 160  
DB 295 -VHERRALIPAVVNCDDLTGSGSEATGRTIARFLGNTSFQG-----RTGAVVVTGSSQ 346

QY 161 TRFTSP-QYVRMRMGDPATPPASVPGTAP-----VPAARAGRTREDVVLASSSQDPTW 213  
DB 347 VHSVRHFKVWSLRDPLGAPAMATVGSWQDGLDFQGAALR-----VPSPGTQARP 400  
QY 214 RLRYDT--SHPTLFORPNH-----VPGMLLY-----BAAR--QAACLYTGPAP 253  
DB 401 KLRVVTIVHHPFVTRFSDSDGQCPAQQLCLDPGTNDSALDIALFALVVGVSVP 454

RESULT 15

US-10-222-772-2  
; Sequence 2, Application US/10222772  
; Publication No. US2004003500A1  
; GENERAL INFORMATION:  
; APPLICANT: Lipton, Stuart A.  
; APPLICANT: Zhang, Dongxian  
; APPLICANT: Chatterton, Jon E.  
; APPLICANT: Awobuluyi, Marc  
; APPLICANT: Sevarino, Kevin A.  
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS  
; FILE REFERENCE: P-LJ 5338  
; CURRENT APPLICATION NUMBER: US/10/222,772  
; CURRENT FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: US 09/934,070  
; PRIOR FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 987  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-10-222-772-2

Query Match 6.2%; Score 101.5; DB 15; Length 987;  
Best Local Similarity 23.8%; Pred. No. 1.6;  
Matches 70; Conservative 31; Mismatches 96; Indels 97; Gaps 13;

QY 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTVQDAFPVSWIPKG-----GDRFSVTA 54  
DB 217 VPAAVVLGCGSTARAEVLEAAPPQOWILGTPLPAEALPTTGTPGVALAGETEQHSLEA 276  
QY 55 VLPHDHPFAPVHGRHDP--LIAETLRQAAMLVFHAGYGVGVHFLMTLDYTGCHLDHL 113  
DB 277 VV-----HDMVELVAQALSSMAL----- 294  
QY 114 GVSGEVALEVEVACSQK-----FRGGQPVQGVDMAVRARGRLAATGTAT 160  
DB 295 -VHERRALIPAVVNCDDLTGSGSEATGRTIARFLGNTSFQG-----RTGAVVVTGSSQ 346  
QY 161 TRFTSP-QYVRMRMGDPATPPASVPGTAP-----VPAARAGRTREDVVLASSSQDPTW 213  
DB 347 VHSVRHFKVWSLRDPLGAPAMATVGSWQDGLDFQGAALR-----VPSPGTQARP 400  
QY 214 RLRYDT--SHPTLFORPNH-----VPGMLLY-----BAAR--QAACLYTGPAP 253  
DB 401 KLRVVTIVHHPFVTRFSDSDGQCPAQQLCLDPGTNDSALDIALFALVVGVSVP 454

Search completed: July 20, 2005, 22:42:56  
Job time : 1166 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: July 20, 2005, 16:48:47 ; Search time 120 Seconds  
(without alignments)  
1008,800 Million cell updates/sec

Title: US-10-017-471B-17  
Perfect score: 1645  
Sequence: 1 MPRAVYLINSASDANSIEQT.....GHQDGLVFLTLTSGPAFSG 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1645	100.0	313	5	ABB80941 S. coelic
2	1645	100.0	313	7	ADE14792 Streptomy
3	524.5	31.9	331	8	ADH39754 Streptomy
4	288	17.5	57	5	ABB80938 S. coelic
5	271.5	16.5	353	4	AAU04042 Streptoco
6	126	7.7	541	7	ADCC31753 Human nov
7	118.5	7.2	516	7	ADCC31754 Human nov
8	118.5	7.2	578	8	ABM80427 Tumour-as
9	118.5	7.2	648	8	ADRI4454 Human NF-
10	110	6.7	749	7	ABO72006 Pseudomon
11	104	6.3	660	7	ABO84125 Pseudomon
12	103	6.3	816	7	ABO74546 Pseudomon
13	101.5	6.2	987	7	ADCC72286 Rat NR3B
14	101.5	6.2	987	8	ADCC72229 Rat NR3B
15	101.5	6.2	987	8	ADN49237 Rat N-met
16	101.5	6.2	987	8	ADN49180 Rat N-met
17	101.5	6.2	1002	7	ADCC72284 Rat NR3B
18	101.5	6.2	1002	7	ADCC72231 Rat NR3B
19	101.5	6.2	1002	8	ADN49182 Rat N-met
20	101.5	6.2	1002	8	ADN49235 Rat N-met
21	101	6.1	250	4	AAU63992 Propionib
22	101	6.1	250	6	ABM60511 Propionib
23	100	6.1	1257	7	ABO68544 Pseudomon
24	99.5	6.0	632	7	ABO68402 Pseudomon
25	99	6.0	1142	4	ABG24027 Novel hum

26	99	6.0	3588	8	ADQ39441 Human myo
27	99	6.0	4346	8	ADQ39440 Human myo
28	99	6.0	4347	8	ADQ39439 Human myo
29	99	6.0	4370	7	ADJ69461 Human hea
30	98.5	6.0	507	5	ADRA1466 Human CD-
31	98.5	6.0	1131	4	ABG21347 Novel hum
32	98	6.0	431	4	ABG20554 Novel hum
33	97.5	5.9	388	7	ABO69171 Pseudomon
34	97	5.9	4391	6	AAE34390 Human per
35	97	5.9	4393	4	AAE31889 Amino aci
36	97	5.9	4393	8	ADL35758 Human per
37	97	5.9	4393	8	ADQ39442 Human myo
38	97	5.9	4336	4	ABG23265 Novel hum
39	96.5	5.9	495	5	ABG66664 Human alp
40	96.5	5.9	495	5	ABG64931 Human alb
41	96.5	5.9	495	5	ABG77002 Human pro
42	96.5	5.9	495	6	ABM84664 Human SEC
43	96.5	5.9	495	8	ADL78198 Albumin f
44	96	5.8	4881	3	ABE23751 S. averm
45	96	5.8	4881	4	AAE5266 Streptomy

## ALIGNMENTS

RESULT 1  
ID ABB80941 standard; protein; 313 AA.

AC ABB80941;

DT 21-OCT-2002 (first entry)

DE S. coelicolor scba protein.

KW Antibiotic; bacterium; scba; afra; scbr; arpa; bara; actinorhodine; Act;  
KW undecylprodigiosin; Red.

OS Streptomyces coelicolor.

PN CA2322241-A1.

PD 23-APR-2002.

PF 23-OCT-2000; 2000CA-02322241.

PR 23-OCT-2000; 2000CA-02322241.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Takano E, Bibb M;

DR WPI, 2002-501089/54.

XX N-PSDB; ABN6379.

PT Modifying antibiotic-producing Streptomyces, to increase, or alter timing  
of, antibiotic production, by deleting the scba or scbr genes.

PS Claim 19; Fig 10; 64pp; English.

XX The invention provides a method for modifying an antibiotic-producing  
strain of Streptomyces to increase production of antibiotics or to alter  
the timing of antibiotic production. The modification is functional

CC deletion of the scba gene of S. coelicolor, or its homologues, but is not

CC deletion of the afra gene of S. griseus, or the modification is

CC functional deletion of the scbr gene of S. coelicolor, or its homologues,  
but is not deletion of arpa of S. griseus nor bara of S. virginiae. The

CC method is particularly used for production of the antibiotics  
actinorhodine (Act) and undecylprodigiosin (Red). The present sequence  
represents the S. coelicolor scba protein

XX Sequence 313 AA;

SQ

Query Match 100.0%; Score 1645; DB 5; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-159; Mismatches 0; Indels 0; Gaps 0;  
 Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTTRVODAFPVSWIPKGGDRFSVTAVLPDH 60  
 DB 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTTRVODAFPVSWIPKGGDRFSVTAVLPDH 60  
 QY 61 PEFAPVHGDRHDPILLIAETLRQAAMLVFHAGYGVGVGHFLMTLDYTCHELDHGVSGEVA 120  
 DB 61 PEFAPVHGDRHDPILLIAETLRQAAMLVFHAGYGVGVGHFLMTLDYTCHELDHGVSGEVA 120  
 QY 121 ELEEVVACSQLKFRGGQPVQGVQVDMVAVRAGRLAATGATTRFTSPQYRRMRGDPATPT 180  
 DB 121 ELEEVVACSQLKFRGGQPVQGVQVDMVAVRAGRLAATGATTRFTSPQYRRMRGDPATPT 180  
 QY 181 ASVPGTAPVPAARAGRTREDEVDVLSASSQDPTWRLRVDTSHPTLQRPNDHVPQMLLLEA 240  
 DB 181 ASVPGTAPVPAARAGRTREDEVDVLSASSQDPTWRLRVDTSHPTLQRPNDHVPQMLLLEA 240  
 QY 241 AROAACLVTPGAPFPVPSIGTRFVRVYAEFDSPCWIOATVRPGPAAGLTTVRVTGHQDGL 300  
 DB 241 AROAACLVTPGAPFPVPSIGTRFVRVYAEFDSPCWIOATVRPGPAAGLTTVRVTGHQDGL 300  
 QY 301 VFLLTTLGGPAPFSG 313  
 DB 301 VFLLTTLGGPAPFSG 313

RESULT 2  
 ADEI4792  
 ID ADEI4792 standard; protein; 313 AA.  
 XX ADEI4792;  
 XX 29-JAN-2004 (first entry)  
 DE Streptomyces coelicolor ScdA protein.  
 XX  
 KW antibiotic-producing strain; antibiotic production; scdA gene; ScdR gene;  
 KM actinorhodin; undecylprodigiosin.  
 XX  
 OS Streptomyces coelicolor.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 102  
 FT /note= "Encoded by ATGGCC on the complementary strand of  
 FT the sequence in figure 14"

XX US2003124644-A1.  
 XX  
 XX 03-JUL-2003.  
 XX  
 XX 23-OCT-2001; 2001US-00017471.  
 XX  
 XX 23-OCT-2000; 2000US-0242561P.  
 XX  
 XX 23-OCT-2000; 2000US-0242561P.  
 XX  
 XX (TAKA/) TAKANO E.  
 XX (BIBB/) BIBB M J.  
 XX  
 XX Takano E, Bibb MJ;  
 XX  
 XX WPI; 2003-810983/76.  
 XX N-PSDB; ADEI4790, ADEI4794.  
 XX  
 XX Modifying an antibiotic-producing strain of Streptomyces coelicolor or  
 XX Streptomyces lividans to increase or alter the timing of antibiotic  
 XX production in the strain, comprises functionally deleting in the strain  
 XX the scdA or ScdR gene.  
 XX  
 XX Claim 19; Fig 10; 33pp; English.  
 XX  
 XX This invention relates to the novel modification of an antibiotic-

CC producing strain of Streptomyces coelicolor or Streptomyces lividans to  
 CC increase or to alter the timing of antibiotic production in the strain.  
 CC The method comprises functionally deleting in the strain the scdA or ScdR  
 CC gene. The method is useful in increasing and altering the timing of  
 CC antibiotic production (especially actinorhodin and undecylprodigiosin) in  
 CC Streptomyces species, particularly Streptomyces coelicolor or  
 CC Streptomyces lividans. The present sequence is that of the Streptomyces  
 CC coelicolor ScdA protein which is used in the method of the invention.  
 CC  
 SQ Sequence 313 AA;  
 XX

Query Match 100.0%; Score 1645; DB 7; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-159; Mismatches 0; Indels 0; Gaps 0;  
 Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTTRVODAFPVSWIPKGGDRFSVTAVLPDH 60  
 DB 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTTRVODAFPVSWIPKGGDRFSVTAVLPDH 60  
 QY 61 PEFAPVHGDRHDPILLIAETLRQAAMLVFHAGYGVGVGHFLMTLDYTCHELDHGVSGEVA 120  
 DB 61 PEFAPVHGDRHDPILLIAETLRQAAMLVFHAGYGVGVGHFLMTLDYTCHELDHGVSGEVA 120  
 QY 121 ELEEVVACSQLKFRGGQPVQGVQVDMVAVRAGRLAATGATTRFTSPQYRRMRGDPATPT 180  
 DB 121 ELEEVVACSQLKFRGGQPVQGVQVDMVAVRAGRLAATGATTRFTSPQYRRMRGDPATPT 180  
 QY 181 ASVPGTAPVPAARAGRTREDEVDVLSASSQDPTWRLRVDTSHPTLQRPNDHVPQMLLLEA 240  
 DB 181 ASVPGTAPVPAARAGRTREDEVDVLSASSQDPTWRLRVDTSHPTLQRPNDHVPQMLLLEA 240  
 QY 241 AROAACLVTPGAPFPVPSIGTRFVRVYAEFDSPCWIOATVRPGPAAGLTTVRVTGHQDGL 300  
 DB 241 AROAACLVTPGAPFPVPSIGTRFVRVYAEFDSPCWIOATVRPGPAAGLTTVRVTGHQDGL 300  
 QY 301 VFLLTTLGGPAPFSG 313  
 DB 301 VFLLTTLGGPAPFSG 313

RESULT 3  
 ADH39754  
 ID ADH39754 standard; protein; 331 AA.  
 XX ADH39754;  
 XX  
 XX 11-MAR-2004 (first entry)  
 DE Streptomyces rochei ORF85 protein SEQ ID NO: 85.  
 XX  
 KW actinomycetes-originated plasmid; linear plasmid; pSLA2-L;  
 KM Streptomyces rochei; secondary metabolite; antibiotic; vitamin;  
 KW lankamycin; mithramycin-like substance; carotenoid; lankacidin.  
 XX  
 OS Streptomyces rochei.  
 XX  
 PN WO2004001039-A1.  
 XX  
 XX 31-DEC-2003.  
 XX  
 XX 19-JUN-2003; 2003WO-JP007767.  
 XX  
 XX 20-JUN-2002; 2002JP-00179345.  
 XX  
 XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 XX Kinashi H;  
 XX  
 XX WPI; 2004-082503/08.  
 XX  
 XX Actinomycetes-originated plasmids for producing secondary metabolites e.g.  
 XX antibiotics and vitamins like lankamycins, lankacidins, mithramycin-like  
 XX substances and carotenoids.

XX Claim 1; SEQ ID NO 85; 499pp; Japanese.  
 PS  
 CC The present invention describes an actinomycetes-originated plasmid is a  
 CC linear plasmid (pSLA2-L) originating in Streptomyces rochei, which has a  
 CC base sequence capable of providing any of the protein sequences of SEQ ID  
 CC No.1 to 143 (ADH39670 to ADH39812), or one based on these sequences but  
 CC with some amino acids deleted, substituted or added and capable of  
 CC producing a secondary metabolite. Also described is a process for  
 CC producing a secondary metabolite by using the plasmid. The plasmid can be  
 CC used for producing secondary metabolites e.g. antibiotics and vitamins  
 CC like lankamycin, mitramycin-like substances, carotenoids and  
 CC lankacidins. Such secondary metabolites are produced selectively and  
 CC efficiently. The present sequence represents a Streptomyces rochei  
 CC protein.  
 CC  
 SQ Sequence 331 AA;  
 Query Match 31.9%; Score 524.5; DB 8; Length 331;  
 Best Local Similarity 38.4%; Pred. No. 1.2e-44;  
 Matches 118; Conservative 44; Mismatches 136; Indels 9; Gaps 5;  
 Oy 10 SASDANSIEQTLAPVPMALVHRTVQDAFPVSWIPKGDRESVAVLPHPDPFAPVHGD 69  
 Db 23 SPGDTSWLPTLPTTVPREYHRASLAEVFLTRCTRIHETRLTGQWFRATTFPLSPDGR 82  
 Oy 70 RHDPLLIAETLRQAMLVFHAAGYGVGVGHFLM-TLDYTCIHDHGVSGEVALEEVAC 128  
 Db 83 RHDPQIAETLRQVGLHAAHAFDVFLGHFIMDMDSFVSRYEHLGVGRTTDLDEATC 142  
 Oy 129 SOLKFRGGQPYQGVDMANVRBAGLAATGTATRT--SPQVYRRMGDFATPTASVPT 186  
 Db 143 VDVVRRRGKLVFPRVITIERDGHVANGG--RTCTTEAMYRLRSAPATTAAHQAAS 200  
 Oy 187 ---APVPAARAGRTREDEVLTASASQDTRLRVDTSHTPLPQBNHVPGLLEBARQ 243  
 Db 201 HQPAPLPSPDGRTPARDVLAFCGAPRWRILNADTSHIIFDHEGDHVPQGVLLSARQ 260  
 Oy 244 AACLVTPG-APFVPSIGSTRFVRYAEFSPCMIOATVAPGAAGLTTVAVTGHODGSLVF 302  
 Db 261 AACALLPFGSTLIPATVSTFRRYVEFTSPCMIEASGLAVTGSCTHALITGRDDDEVF 320  
 Oy 303 LTTLSGP 309  
 Db 321 TARISGP 327  
 RESULT 4  
 ID ABB80938 standard; protein; 57 AA.  
 AC ABB80938;  
 DT 21-OCT-2002 (first entry)  
 DE S. coelicolor scdA protein fragment.  
 XX Antibiotic; bacterium; scdA; afsA; scdR; arpA; barA; actinorhodine; Act;  
 KM undecylprodigiosin; Red.  
 XX Streptomyces coelicolor.  
 OS CA2322241-A1.  
 FN CA2322241-A1.  
 PD 23-APR-2002.  
 PF 23-OCT-2000; 2000CA-02322241.  
 PR 23-OCT-2000; 2000CA-02322241.  
 PA (PLAN-) PLANT BIOSCIENCE LTD.  
 PI Takano E, Bibb M;

XX WPI; 2002-501089/54.  
 DR N-PSDB; ABN86377.  
 XX  
 PT Modifying antibiotic-producing Streptomyces, to increase, or alter timing  
 PT of, antibiotic production, by deleting the scdA or scdR genes.  
 XX  
 PS Disclosure; Fig 4B; 64pp; English.  
 CC The invention provides a method for modifying an antibiotic-producing  
 CC strain of Streptomyces to increase production of antibiotics or to alter  
 CC the timing of antibiotic production. The modification is functional  
 CC deletion of the scdA gene of S. coelicolor, or its homologues, but is not  
 CC deletion of the afsA gene of S. griseus, or the modification is  
 CC functional deletion of the scdR gene of S. coelicolor, or its homologues,  
 CC but is not deletion of arpA of S. griseus nor barA of S. virginiae. The  
 CC method is particularly used for production of the antibiotics  
 CC actinorhodine (Act) and undecylprodigiosin (Red). The present sequence  
 CC represents the S. coelicolor scdA protein fragment  
 SQ Sequence 57 AA;  
 Query Match 17.5%; Score 288; DB 5; Length 57;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-21;  
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 MPEAVVLINSASDANSIEQTLAPVPMALVHRTVQDAFPVSWIPKGDRESVAVLP 57  
 Db 1 MPEAVVLINSASDANSIEQTLAPVPMALVHRTVQDAFPVSWIPKGDRESVAVLP 57  
 RESULT 5  
 ID AAU04042 standard; protein; 353 AA.  
 AC AAU04042;  
 DT 11-SEP-2003 (revised)  
 DT 23-OCT-2001 (first entry)  
 DE Streptococcus coelicolor Mmfl protein.  
 XX SCP1; methylenomycin cluster; mmc; MmyR; Mmfl; Mmfl; Mmfl; Mmfl;  
 KM MmyO; MmyQ; MmyJ; Mmr; heterologous gene expression.  
 XX Streptomyces coelicolor.  
 OS WO200148228-A1.  
 FN WO200148228-A1.  
 PD 05-JUL-2001.  
 PF 20-DEC-2000; 2000WO-GB004972.  
 PR 23-DEC-1999; 99GB-00030477.  
 PA (PLAN-) PLANT BIOSCIENCE LTD.  
 PI Chapter KF, Bruton CJ, O'rouke SJ, Wietzorrek AW;  
 DR WPI; 2001-425675/45.  
 DR N-PSDB; AAS07627, AAU04045, AAU04046.  
 PT Novel expression cassette for expressing a nucleic acid of interest,  
 PT derived from the regulatory region of methylenomycin gene cluster of SCP1  
 PT plasmid of Streptomyces coelicolor A3(2).  
 XX  
 PS Claim 25; Fig 8d; 142pp; English.  
 CC The sequence represents the Mmfl protein encoded by the mmfl gene carried  
 CC on the expression cassette present on plasmid SCP1. The expression  
 CC cassette is the regulatory region of the methylenomycin cluster (mmc)  
 CC from Streptomyces coelicolor A3(2), which encodes the MmyR, Mmfl, Mmfl,  
 CC Mmfl, Mmfl, MmyO, MmyG, MmyJ and partial Mmr polypeptides. The

CC expression cassette is useful for expressing a nucleic acid of interest,  
 CC substantially only when the host cell culture reaches high cell density  
 CC at or close to the stationary phase of host cell culture. In particular  
 CC the system is useful in regulating methylenomycin production. Reduced or  
 CC no expression of the nucleic acid of interest is observed earlier in  
 CC growth, avoiding toxic effects of some gene products on growth and the  
 CC system does not require addition of exogenous inducer. The methylenomycin  
 CC cluster naturally present on a highly transmissible plasmid permits  
 CC properly regulated expression in diverse Streptomyces host and the  
 CC expression is driven by a strong promoter, leading to high yield of the  
 CC desired end product. (Updated on 11-SEP-2003 to standardise OS field)

XX  
 SQ Sequence 353 AA;

Query Match 16.5%; Score 271.5; DB 4; Length 353;  
 Best Local Similarity 27.9%; Pred. No. 1e-18;  
 Matches 95; Conservative 43; Mismatches 123; Indels 79; Gaps 12;

QY 21 ALPVMALVHRTVODAFVSVIPIKGGDFSTYAVLPHDHPFAVHGDRHPLLAETL 80  
 DB 26 ARLPPADVHKAAMAEVLLTDARPIGENRFAVAALMPRTFLAHRATSSPCDPLLAETI 85  
 QY 81 RQAMLVFAGYGVPGVYFLMTLDYTLCHLHGVSEVALEVEVACSGQLKFRGGQPVQ 140  
 DB 86 RQSAHLHSTFCDDVPIGHIFVL-----SGLDLDLDLVWDS-----GPLPV 127  
 QY 141 GQVDMVVRBAGLIAATGTAAT-----RP--TSPQVY--- 169  
 DB 128 -----LDVTSTKTTNPRRMAPALNADYVAGLHGRCAIREVLAPRRVAMI 175  
 QY 170 --RRRRGDPATPTASVPGTAPPARAGTRDEVDVL--SASSQOFT--WLRVDTSHTPLF 225  
 DB 176 RDRARRAEKPPAQOAAAGATATLPEETVGHDDHLVLAHQGLPTAQWLKRRDHPVLF 235  
 QY 226 QRPNDHVPMLLLEAARQAACVTGPAP--FVP-----SIGTRFVRYAEFSPCMIOAT 278  
 DB 236 DHESHISMALLECRQAATLTTPAPCAFSPROVALTAVASSVOAREGLDSP--VIT 293  
 QY 279 VAP-----GPAAGLTVTVYGHQDGSVLVPLTTLGPAFSG 313  
 DB 294 TLPAAHGHSPPDSGTRTQLQTARQGSRTLTATVTTTGTG 333

RESULT 6  
 ADC31753  
 ID ADC31753 standard; protein; 541 AA.

AC ADC31753;

XX 18-DEC-2003 (first entry)

DE Human novel polypeptide sequence, SEQ ID NO:1835.

XX Human; diagnostic; drug screening; forensics; gene mapping;  
 KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
 KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
 KW ulcers; osteoporosis; autoimmune disease; cancer;  
 KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
 KW neuroprotective; anti-anemic; anticoagulant; thrombolytic; vulnerary;  
 KW antitumor; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
 KW gene therapy; chromosome 3.

XX Homo sapiens.

XX WO2003029271-A2.

XX 10-APR-2003.

XX 24-SEP-2002; 2002WO-US030474.

XX 24-SEP-2001; 2001US-0324631P.

XX (HYSE-) HYSEQ INC.

XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
 PI Haley-Vicente D, Dmanac RT;  
 DR MPI: 2003-371981/35.  
 DR N-P8DB; ADC30782.

XX New polynucleotide and polypeptide useful for diagnosing, preventing or  
 PT treating conditions such as neurodegenerative diseases, anemias, platelet  
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 PT cancer.

PS Claim 20; SEQ ID NO 1835; 1185bp; English.

XX The invention relates to 971 novel human cDNA sequences (ADC29919-  
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
 CC invention also relates to nucleic acid sequences over 99% identical with  
 CC the novel human cDNAs. The invention additionally encompasses expression  
 CC vectors and host cells comprising a nucleic acid of the invention; the  
 CC recombinant production of a polypeptide of the invention; an antibody  
 CC against a polypeptide of the invention; a method of detecting  
 CC polynucleotides or polypeptides of the invention; and methods of  
 CC identifying a compound which binds to a polypeptide of the invention. The  
 CC invention further discloses methods of preventing, treating or  
 CC ameliorating a medical condition; kits comprising polynucleotide probes  
 CC and/or monoclonal antibodies for carrying out the methods of the  
 CC invention; methods for the identification of compounds that modulate the  
 CC expression or activity of the polynucleotide and/or polypeptide; and 767  
 CC coding sequences corresponding to the cDNA sequences of the invention  
 CC (ADC31861-ADC32627) and the polypeptides encoded by the cDNAs (ADC32628  
 CC -ADC33394). The nucleic acids and polypeptides of the invention are  
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
 CC identification of mutations responsible for genetic disorders or other  
 CC traits, for assessing biodiversity, and in producing many other types of  
 CC data and products dependent on DNA and amino acid sequences. They are  
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, anaemia, platelet  
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 CC cancer. The nucleic acids may also be used as hybridisation probes or  
 CC primers, and in the recombinant production of a protein. The polypeptides  
 CC are also useful in generating antibodies, as molecular weight markers,  
 CC and as food supplements. The present sequence represents a specifically  
 CC claimed human polypeptide sequence of the invention. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 541 AA;

Query Match 7.7%; Score 126; DB 7; Length 541;  
 Best Local Similarity 28.6%; Pred. No. 0.0014;  
 Matches 55; Conservative 21; Mismatches 90; Indels 26; Gaps 7;

QY 50 FSVAVLPHDHPFAPVHGDRHDL-----LAEITRQAAMLVFNHGVGVYHFLM 102

DB 77 FSLAL-----HPVQPTHERKPTLEHIFRAILLIKKIDRSQKTDLHNGY-----IL 124

QY 103 TLDTYCHLDHGVSEVAEL--EVEVACSG--LKFRGGQPGVGVDMVARRAGRLAATGTA 159

DB 125 ELDDCCSLDHPDQKLIPEFTIKIQEAAAGQLKFGVLP---QTHSSVNSAGSSAPVSTA 181

QY 160 TTRFTSPGVYRRMRGDFATPTASVGTAPVAPARAGTRDEVDVLASASSQDTRWLRVDT 219

DB 182 NSTEDARDA-KVAGDHDASLENEKGTDCVAPAGRNQSPSPSGPRGVEHEPLSGKM 240

QY 220 SHPTLFGRPNDH 231

DB 241 EIFTLENPKSH 252

RESULT 7  
 ADC31754

ID ADC31754 standard; protein; 516 AA.  
 AC ADC31754;  
 DT 18-DEC-2003 (first entry)  
 DE Human novel polypeptide sequence, SEQ ID NO:1836.  
 XX  
 XX Human; diagnostic; drug screening; forensics; gene mapping;  
 KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
 KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
 KW ulcers; osteoporosis; autoimmune disease; cancer;  
 KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
 KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnary;  
 KW antifuror; osteopathic; immunosuppressive; antiinflammatory; cyostatic;  
 KW gene therapy; chromosome 3.  
 XX  
 OS Homo sapiens.  
 PN WO2003029271-A2.  
 XX  
 XX 10-APR-2003.  
 PD 24-SEP-2002; 2002WO-US030474.  
 PF 24-SEP-2001; 2001US-0324631P.  
 PR 24-SEP-2001; 2001US-0324631P.  
 PA (HYSE-) HYSEQ INC.  
 PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehman T;  
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
 PI Haley-Vicente D, Drmanac RT;  
 XX  
 DR WPI; 2003-371981/35.  
 DR N-PSDB; ADC30783.  
 XX  
 PT New polynucleotide and polypeptide useful for diagnosing, preventing or  
 PT treating conditions such as neurodegenerative diseases, anemias, platelet  
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 PT cancer.  
 PS Claim 20; SEQ ID NO 1836; 1185bp; English.  
 XX  
 CC The invention relates to 971 novel human cDNA sequences (ADC29919-  
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
 CC invention also relates to nucleic acid sequences over 99% identical with  
 CC the novel human cDNAs. The invention additionally encompasses expression  
 CC vectors and host cells comprising a nucleic acid of the invention; the  
 CC recombinant production of a polypeptide of the invention; an antibody  
 CC against a polypeptide of the invention; a method of detecting  
 CC polynucleotides or polypeptides of the invention; and methods of  
 CC identifying a compound which binds to a polypeptide of the invention. The  
 CC invention further discloses methods of preventing, treating or  
 CC ameliorating a medical condition; kits comprising polynucleotide probes  
 CC and/or monoclonal antibodies for carrying out the methods of the  
 CC invention; methods for the identification of compounds that modulate the  
 CC expression or activity of the polynucleotide and/or polypeptide; and 767  
 CC coding sequences corresponding to the cDNA sequences of the invention  
 CC (ADC31861-ADC32627) and the polypeptides encoded by the cDNAs (ADC32628  
 CC -ADC33394). The nucleic acids and polypeptides of the invention are  
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
 CC identification of mutations responsible for genetic disorders or other  
 CC traits, for assessing biodiversity, and in producing many other types of  
 CC data and products dependent on DNA and amino acid sequences. They are  
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 CC cancer. The nucleic acids may also be used as hybridisation probes or  
 CC primers, and in the recombinant production of a protein. The polypeptides  
 CC are also useful in generating antibodies, as molecular weight markers,  
 CC and as food supplements. The present sequence represents a specifically  
 CC claimed human polypeptide sequence of the invention. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but

CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 516 AA;  
 Query Match 7.2%; Score 118.5; DB 7; Length 516;  
 Best Local Similarity 27.1%; Pred. No. 0.0077;  
 Matches 62; Conservative 22; Mismatches 82; Indels 63; Gaps 11;  
 QY 50 FSVYAVLPHPHPFAPVGRHDPDLP-----LIAETLRQAMLVFAGYGVPGVYHFLM 102  
 DB 77 FSLAAL-----HPFQPHREKPTLPHIFRAILIKTRDSQKTDLHNEG-----IL 124  
 QY 103 TLDYTCCHDLGVSGEVAEL--EVEVACSQ-LKFRGQPVQGVQVMAVPRAGRLATGTA 159  
 DB 125 ELDCSSLDHPDQKILPEFTIKKIQAASQGLKFGVGP---QYHSSVNSAGSSAPVSTA 181  
 QY 160 TTRFTSPQVTRRMAGDAPATPTASVPGTAPYPPARAAGTRDE-----DVL-----SA 206  
 DB 182 NSTEDARDA-KNMGDHASDENKPGTGVCAPAGNQSPEPSSGPRGEVPLAKQSSP 240  
 QY 207 SSGQDTRLRVD-----TSHP-----TLFORPMDH 231  
 DB 241 SSGGDGGLSPQGVSKTLDGPESNPLEVHEEPLSGKMEITTLPRKPSH 289  
 RESULT 8  
 ABM80427  
 ID ABM80427 standard; protein; 578 AA.  
 AC ABM80427;  
 XX  
 DT 18-NOV-2004 (first entry)  
 DE Tumour-associated antigenic target (TAT) polypeptide PRO80962, SEQ:1073.  
 XX  
 KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
 KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
 KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
 KW central nervous system cancer; bladder cancer; pancreatic cancer;  
 KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
 KW chromosome identification; chromosome mapping; gene mapping;  
 KW gene therapy; cyostatic.  
 XX  
 OS Homo sapiens.  
 PN WO2004030615-A2.  
 EN 15-APR-2004.  
 PD 29-SEP-2003; 2003WO-US028547.  
 PF 02-OCT-2002; 2002US-0414971P.  
 PR (GENTH ) GENENTECH INC.  
 PA Wu TD, Zhang Z, Zhou Y;  
 PI WPI; 2004-347921/32.  
 DR N-PSDB; ACN37889.  
 XX  
 PT New tumor-associated antigenic target polypeptides and nucleic acids,  
 PT useful in preparing a medicament for treating or detecting a  
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
 PT prostate cancer or tumor.  
 PS Claim 12; SEQ ID NO 1073; 7273bp; English.  
 XX  
 CC The invention relates to human tumour-associated antigenic target (TAT)  
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
 CC overexpressed in cancer tissues compared to normal tissues, and may thus  
 CC serve as effective targets for the diagnosis and treatment of cancer in  
 CC mammals. The invention also relates to nucleic acid and polypeptide





AC ABO72006;  
XX 29-JUL-2004 (first entry)  
XX Pseudomonas aeruginosa polypeptide #4181.  
XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
XX Pseudomonas aeruginosa.  
XX US6551795-B1.  
XX 22-APR-2003.  
XX 18-FEB-1999; 99US-00252991.  
XX 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX WPI; 2003-615309/58.  
XX N-PSDB; ABD05577.  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathologial conditions resulting from bacterial infection.  
XX Disclosure; SEQ ID NO 20752; 455pp; English.  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnostics and  
XX therapy of pathologial conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathologial conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences ABO67826-  
XX ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX seqdata.uspto.gov/sequence.html  
SQ Sequence 749 AA;  
Query Match 6.7%; Score 110; DB 7; Length 749;  
Best Local Similarity 24.7%; Pred. No. 0.097;  
Matches 74; Conservative 26; Mismatches 116; Indels 84; Gaps 15;  
43 IPKGGDRSVTAVALPHDHPFAPVYGDHDDLLIA-----ETLRQAMLVFHAQYVGP 95  
465 VPEREDLAGMNVLAGHGLDAGAGEGHDALVAGFQQLLDITGDLPGHIFGAG-ARP 523  
66 VGYHFLMTLDYTCMLDLGVSGEVALEEV-EVACSQQLKFRGGQPVQGVDMAV--RRGR 152  
524 VGLH-----HHGLSEGERVFPALPOLAVGQRADRAKQHDQVDDLAIIQRPFGK 571  
153 LAATGTATT-----RTSPQVYRRMRGDFATPTASVPG--TAEPVPAARAGTR 198  
572 V--EGHATTMPWPVPSPCGMPGRSDAPALREAPGRMPPSRQPSGLRAPAPGP--RHR 626  
199 DEDVVLASASQODTRRLRVDTSHPTLFFORPNDHVPGMLLLEAARQAACLVMP----- 251  
667 DRCAPPAAG-----SSCPA--RPPARTTGRPLRSHPTREP--ITAPPPRRRRR 671  
252 ----APFVPSIGSTRFV-----RYAEFDS-----CWIQATVRRPAPAGLITTVR 291

DB 672 PARSGVPRESRGRRRPPGRRRYALADRRPARARAGVARRSVCRAVTRVPTPAGRCGNVR 731  
RESULT 11  
ID ABO84125  
ID ABO84125 standard; protein; 660 AA.  
XX ABO84125;  
XX 29-JUL-2004 (first entry)  
XX Pseudomonas aeruginosa polypeptide #16300.  
XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
XX Pseudomonas aeruginosa.  
XX US6551795-B1.  
XX 22-APR-2003.  
XX 18-FEB-1999; 99US-00252991.  
XX 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX WPI; 2003-615309/58.  
XX N-PSDB; ABD17696.  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathologial conditions resulting from bacterial infection.  
XX Disclosure; SEQ ID NO 32871; 455pp; English.  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnostics and  
XX therapy of pathologial conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathologial conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences ABO67826-  
XX ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX seqdata.uspto.gov/sequence.html  
SQ Sequence 660 AA;  
Query Match 6.3%; Score 104; DB 7; Length 660;  
Best Local Similarity 25.9%; Pred. No. 0.33;  
Matches 72; Conservative 17; Mismatches 131; Indels 58; Gaps 12;  
34 VQDAFPVSWIPKGGDRSVTAVALPHDHPFAPVYGDHDDLLIAETLRQAMLVFHAQY 92  
1 IPSAMPPSMBRCCMPR-----RPSDDP-----GVN-----PLQILPLRRPAPLHAHPRGM 46  
93 GVPVGYHFLMTLDY--TCMLDLGVSGEVALEEVACSQQLKFRGGQPVQGVDMAV--RRA 150  
47 SAARARHRLYADIGSGRRAGRLSAGRGALDIDQPRAGRVHNVGRRPRQADAPGADRL 106  
151 GRILATGTATRTFTSPQVYRRMRGDFATPTASVPGTAVPDAARAGTRDEDVVLASASQO 210  
107 QODARTGCLLR--RRLFRGKGPDLAQP--AVPVPGLPLPLR----- 144



```

Qy      114 GVSQGEVLEVEVAVCSQK-----FRGGQPVGGQVDMWVRBAGRLAATGTAT 160
Db      295 -VHPRALLPAVAVCDLKTGSGSEATGRTLARELGNTISFGQ-----RTGAWVVTGSSQ 346
Qy      161 TRFTSP-QVYRRMGDFATPTASVGTAP-----VPARAGRTREDDVLSASSQDPTW 213
Db      347 VHVSHRFVWMLRNDPLGAPAMATVGSWGQGLDFQGAALR-----VPSPGTQARP 400
Qy      214 RLRYDT--SHPTLFQRPNDH---VPGMLL-----EAAR---QAACLVTPAP 253
Db      401 KLRVVTLVHPFVFTRESDEGQCPAGQLCLDPGTNDSARLDALFALVNGSV 454

```

## RESULT 14

```

ADCT72229
ID      ADC72229 standard; protein; 987 AA.

```

```

XX      AC      ADC72229;

```

```

XX      DT      18-DEC-2003 (first entry)

```

```

XX      DE      Rat NR3B SEQ ID NO:2.

```

```

XX      rat; NR3B: cerebroprotective; nootropic; antiparkinsonian; gene therapy;

```

```

XX      stroke; Alzheimer's disease; Parkinson's disease; N-methyl-D-aspartate;

```

```

XX      NMDA; N-methyl-D-aspartate receptor type 3B.

```

```

XX      Rattus sp.

```

```

XX      WO2003016479-A2.

```

```

XX      27-FEB-2003.

```

```

XX      16-AUG-2002; 2002WO-US026022.

```

```

XX      20-AUG-2001; 2001US-00934070.

```

```

XX      (BURN-) BURNHAM INST.

```

```

XX      (UYVA ) UNIV YALE.

```

```

XX      (BGHM ) BRIGHAM & WOMENS HOSPITAL.

```

```

XX      Lipton SA, Zhang D, Chatterton JE, Awobuluyi M, Sevarino KA;

```

```

XX      WPI; 2003-278559/27.

```

```

XX      DR      N-PSDB; ADC72228.

```

```

XX      PT      New NR3B nucleic acid molecule encoding a NR3B polypeptide, useful for

```

```

XX      PT      preparing a composition for treating e.g. stroke, Alzheimer's disease or

```

```

XX      PT      Parkinson's disease.

```

```

XX      PS      Claim 21; SEQ ID NO 2; 190pp; English.

```

```

XX      CC      The invention relates to a novel NR3B nucleic acid molecule. A NR3B

```

```

XX      CC      protein has cerebroprotective, nootropic, and antiparkinsonian activity.

```

```

XX      CC      A gene of the invention may have a use in gene therapy. The NR3B nucleic

```

```

XX      CC      acid is useful for preparing a composition for treating e.g. stroke,

```

```

XX      CC      Alzheimer's disease or Parkinson's disease. The present sequence is used

```

```

XX      CC      in the exemplification of the invention.

```

```

XX      SQ      Sequence 987 AA;

```

```

Query Match      6.2%; Score 101.5; DB 7; Length 987;

```

```

Best Local Similarity 23.8%; Pred. No. 1.1;

```

```

Matches 70; Conservative 31; Mismatches 96; Indels 97; Gaps 13;

```

```

Qy      1 MPEAVVLINSASDANSIEQTALPVPMALVHRTVQDAFPVSWIPKG-----GDRFSVTA 54

```

```

Db      217 VPAVVLIGCSTYRAHVEVLEAARPGQWLTGRLPRLAALPTGLRPGVALAGTEQHSIEA 276

```

```

Qy      55 VLPHDHPFAVHGGRHDP--LIAETLEQAMLVFNHAGYGVPGVGHPLMTLDYCHDHL 113

```

```

Db      277 VV-----HDMVELVAVQALSSMAL----- 294

```

```

Qy      114 GVSQGEVLEVEVAVCSQK-----FRGGQPVGGQVDMWVRBAGRLAATGTAT 160
Db      295 -VHPRALLPAVAVCDLKTGSGSEATGRTLARELGNTISFGQ-----RTGAWVVTGSSQ 346
Qy      161 TRFTSP-QVYRRMGDFATPTASVGTAP-----VPARAGRTREDDVLSASSQDPTW 213
Db      347 VHVSHRFVWMLRNDPLGAPAMATVGSWGQGLDFQGAALR-----VPSPGTQARP 400
Qy      214 RLRYDT--SHPTLFQRPNDH---VPGMLL-----EAAR---QAACLVTPAP 253
Db      401 KLRVVTLVHPFVFTRESDEGQCPAGQLCLDPGTNDSARLDALFALVNGSV 454

```

## RESULT 15

```

ADN49237
ID      ADN49237 standard; protein; 987 AA.

```

```

XX      AC      ADN49237;

```

```

XX      DT      01-JUL-2004 (first entry)

```

```

XX      DE      Rat N-methyl-D-aspartate (NMDA) type 3B (NR3B) B4 protein #2.

```

```

XX      N-methyl-D-aspartate; NMDA; NMDA type 3B; NR3B; stroke; ischaemia;

```

```

XX      trauma; headache; epilepsy; neuropathic pain syndrome;

```

```

XX      diabetic neuropathy; glaucoma; depression; anxiety; drug addiction;

```

```

XX      Alzheimer's disease; Huntington's disease; HIV-associated dementia;

```

```

XX      Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;

```

```

XX      gene therapy; rat.

```

```

XX      Rattus sp.

```

```

XX      OS      Location/Qualifiers

```

```

XX      FH      Key

```

```

XX      FT      Peptide

```

```

XX      FT      Protein

```

```

XX      FT      Binding-site

```

```

XX      FT      Region

```

```

XX      FT      Region

```

```

XX      FT      Region

```

```

XX      FT      Binding-site

```

```

XX      FT      Region

```

```

XX      FT      Region

```

```

XX      FT      Region

```

```

XX      FT      Region

```

```

XX      FT      Region

```

```

XX      FT      Region

```

```

XX      FT      Region

```

```

XX      FT      Region

```

```

XX      FT      Region

```

```

XX      FT      Region

```

CC The invention relates to N-methyl-D-aspartate (NMDA) type 3B (NR3B)  
CC protein and its corresponding nucleic acid sequence. The invention is  
CC useful for preventing or treating conditions in which inappropriate NMDA  
CC receptor activation or inappropriate responses to glycine or glutamate  
CC are implicated, such as stroke, ischemia, head and spinal trauma,  
CC headache, epilepsy, neuropathic pain syndromes including diabetic  
CC neuropathy, glaucoma, depression and anxiety, drug addiction, Alzheimer's  
CC disease, Huntington's disease, HIV-associated dementia, Parkinson's  
CC disease, multiple sclerosis or amyotrophic lateral sclerosis. It is also  
CC useful in gene therapy. The present sequence is rat NR3B B4 protein.

XX  
SQ Sequence 987 AA;

Query Match

Best Local Similarity 23.8%; Score 101.5; DB 8; Length 987;

Matches 70; Conservative 31; Mismatches 96; Indels 97; Gaps 13;

QY 1 MEPAVLINASDANSIEQTALVPVMAIVHRTRVQDAPVSWIPKG-----GDRFSVTA 54  
DB 217 VPAAYVLGCGSTAHAEVLEAAPPQWLGTPLPALPTTGLPQGVIALGETEQHSLER 276  
QY 55 VLPHDHPFFAPVHGHDPL-LIAETLRQAMLVFHAGYGVVGYHFLMTLDYTCILDHL 113  
DB 277 VV-----HDVVELVAQALSMAL----- 294  
QY 114 GVSGEVABEVEVACSQK-----FRGGQPVQGVDMVVRPAGRLAATGTAT 160  
DB 295 -VHPERALIPAVVNCDDLKTGSEATGRTIARFLGNTSFQG-----RTGAVWVYGSQ 346  
QY 161 TTFSTP-QYRRMRGDFATPTASVPGTAP-----VPAARAGTRDEDEVLSASSQDTW 213  
DB 347 VHVSRHFKVMSLRDPLGAPAWATVGSWDDGQLDFOPGAALR-----VSPSGTQARP 400  
QY 214 RLKVDL--SHPTLFORPNDH---VPGMLL-----EAAR---QAACLVTPAP 253  
DB 401 KLRVVTLVVHFPVFTRESDEDDGCPAGQCLDPTNDSARLDALPALVNGSVP 454

Search completed: July 20, 2005, 22:10:05  
Job time : 127 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 20:11:08 ; Search time 38 Seconds  
(without alignments)  
792.522 Million cell updates/sec

Title: US-10-017-471B-17

Perfect score: 1645

Sequence: 1 MPEAVVLINSADANSIEQT.....GHQDGLVFLTLGPAFSG 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1023.5	62.2	301	2 A32061	af6a protease - Str
2	109.5	5.7	1762	2 T03222	probable polyketid
3	97	5.9	4391	2 A38096	perlecan precursor
4	96.5	5.9	474	1 OMHUIB	alpha-1-8-glycopro
5	94.5	5.7	323	2 H75580	conserved hypotet
6	93	5.7	362	2 T36079	hypothetical prote
7	93	5.7	559	2 T02825	probable membrane
8	93	5.7	786	2 S22155	oncogene 1 (tre-2
9	91.5	5.6	307	2 C75583	hypothetical prote
10	91.5	5.6	407	2 G84309	hypothetical prote
11	91.5	5.6	472	1 F70876	probable pap3 pro
12	91.5	5.6	1634	2 T26517	hypothetical prote
13	91	5.5	912	2 B97566	hypothetical prote
14	91	5.5	1008	2 AH2786	conserved hypotet
15	90.5	5.5	345	2 AH0350	probable DNA-bindi
16	90	5.5	240	2 S30406	hypothetical prote
17	90	5.5	438	2 T34683	hypothetical prote
18	90	5.5	646	2 T35002	probable respirato
19	89.5	5.4	676	2 AG0035	hemim receptor pre
20	89.5	5.4	676	2 T12069	hemim receptor - y
21	89	5.4	365	2 F96017	hypothetical prote
22	89	5.4	767	2 S41479	DNA-binding protei
23	89	5.4	1460	1 EDBEIR	DNA-binding protei
24	89	5.4	2282	2 T42717	immediate-early pr
25	89	5.4	5149	2 F83345	DNA-binding protei
26	88.5	5.4	607	2 S42639	probable non-ribos
27	88.5	5.4	5069	2 T17464	ATP-dependent RNA
28	88.5	5.4	7463	2 T36248	ribozyme polyketi
29	88	5.3	405	2 T37022	CDA peptide synthe

30	87.5	5.3	172	2 H87253	hypothetical prote
31	87.5	5.3	366	2 C83034	probable oxidoredu
32	87.5	5.3	366	2 T36169	hypothetical prote
33	87.5	5.3	422	2 H75501	DNA/pantothenate m
34	87.5	5.3	626	2 S53871	Pmel 17 protein -
35	87	5.3	559	2 C75286	hypothetical prote
36	87	5.3	606	2 T35710	transferase - stre
37	87	5.3	1159	2 E70741	probable regulator
38	87	5.3	1241	2 T18311	hypothetical prote
39	86.5	5.3	639	2 F95324	trac conjugal tran
40	86	5.2	455	2 D84371	phosphomannomutase
41	86	5.2	625	2 S35317	hematopoietic grow
42	86	5.2	4735	2 T17463	ribozyme polyketi
43	85.5	5.2	279	2 AG3540	dihydrodipicolinat
44	85.5	5.2	315	2 S72887	hydroxymethylbilan
45	85.5	5.2	435	2 T03545	probable cobyrimic

#### ALIGNMENTS

RESULT 1  
af6a protease - Streptomyces griseus  
C/Species: Streptomyces griseus  
C/Date: 31-Aug-1990 #sequence\_revision 31-Aug-1990 #text\_change 09-Jul-2004  
C/Accession: A32061  
R/Horinouchi, S.; Suzuki, H.; Nishiyama, M.; Beppu, T.  
J. Bacteriol. 171, 1206-1210, 1989  
A/Title: Nucleotide sequence and transcriptional analysis of the Streptomyces griseus ger  
A/Reference number: A32061; MUID:89123125; PMID:2492509  
A/Accession: A32061  
A/Status: preliminary  
A/Molecule type: DNA; mRNA  
A/Residues: 1-301 <HOR>  
A/Cross-reference: UNIROF:P18394; GB:M4250; NID:g153148; PID:AAA26693.1; PID:g153149  
C/Superfamily: Streptomyces griseus af6a protease

Query Match 62.2%; Score 1023.5; DB 2; Length 301;  
Best Local Similarity 66.4%; Pred. No. 3.2e-78;  
Matches 194; Conservative 26; Mismatches 71; Indels 1; Gaps 1;

Qy	23	PVPMAVHRTTRVODAFVSWIPKGRFSTAVLPHDHPFAPVGHGHDPLIAETLRQ	82
Db	10	PVGIEVHRTTRPEDAFPMNWRLGRDRFAVEAVLPHDHPFAPVGDPLIAEAMRQ	69
Qy	83	AAAMVFAGYGVPGYHFLMT-IDYTHLDHLGSGVALEEVAVCSQKFRGGQVVG	141
Db	70	AAAMAFAGYGVPGYHFLMTIDYCHPEHLGVGSEPTIGLEVFCSDKMRGLPAQG	129
Qy	142	QVDAVVRAGRIAAATGATTRFTSPQYRRMRGDFATPTASVPGTAVPAPARAGRTDED	201
Db	130	RVGMAVRGRLAATGVAATRFSTPKAYRRMRGDPVEGLSLEPTAVVPASPRARVED	189
Qy	202	VVLSSSQDTRWLKVDTSPTLFQRENDHVPGLILAAARQAACVTGAPAFVPSIGT	261
Db	190	VVLSTGRGWEHLRVDTTRPTLFQRENDHVPGLILAAARQAACVAGAGATPVVART	249
Qy	262	RFRVYAFDGPWCWQATVTRPGPAAGLTTVRYTHGQDSLYFLTLGSPAASG	313
Db	250	RFRYSEFGSPCWIGAVVPGADEDTVTRVTHGQDETFTSLSGPRAHG	301

RESULT 2  
T03222  
Probable polyketide synthase module 2 - Streptomyces hygroscopicus  
C/Species: Streptomyces hygroscopicus  
C/Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
C/Accession: T03222  
R/Ruan, X.; Staehli, D.; Lax, S.; Katz, L.  
Gene 203, 1-9, 1997  
A/Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC 2

A,Accession: T03222  
 A,Status: preliminary; translated from GB/EMBL/DBJ  
 A,Molecule type: DNA  
 A,Residues: 1-1762 <MOR>  
 A,Cross-references: UNIPROT:Q030480; EMBL:AF007101; NID:G2624946; PIDD:AAAC38062.1; PID:92  
 A,Experimental source: ATCC 29253  
 C,Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac  
 homology; [acyl-carrier-protein] S-malonyltransferase homology  
 C,Keywords: carrier protein  
 F:54-454/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>  
 F:550-822/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>  
 F:1558-1669/Domain: acyl carrier protein homology <ACP>

Query Match 6.7%; Score 109.5; DB 2; Length 1762;  
 Best Local Similarity 24.6%; Pred. No. 0.92;  
 Matches 93; Conservative 32; Mismatches 162; Indels 91; Gaps 19;

QY 10 SASDANSLEQTLPLPMLVHRTTRQDAFPVSWIKGDRSVTVAVLHHDHFFAPV--- 66  
 DB 697 SIAMNGPRHATVLSGDRITLHRIATQMLTKTNWLNVSHAFSPMLQPILOFTTTTLNTL 755  
 QY 67 -HGDRHDPILLIAETLRQAAMLVFHAQGVGVGFHMTLDYCHDHLGVSSEVALEEV- 124  
 DB 756 THHPHTPLISMILTPPHPTTH-----WTQHTAPRYTDTLHLLHHCITTYLEIG 809  
 QY 125 -EVACSQLKFRGQGVQGVDAVVR-----AGRLAATG-----TATTFPT 164  
 DB 810 PPTTLTALA-RTTLPTTTHLIPTRRNNEVASTNEALGRVSVGHSDVMRLTPTGRT 868  
 QY 165 SPQVRRNRGDPATPTAAYVGPAPVPAARAGTRBEDVYLSAQDTRRL---RVDT 220  
 DB 869 SLPTPTFGRRDPMHDA-GGGAEEVGAGL-G-TTDPHPLGAVTVADTGLLISRLSTS 926  
 QY 221 -HPTFORNDH-----VPGMLLEAARQA-----CLVTGAPAPVP 256  
 DB 927 THFWL---TDHTVNGTVVPEGTALIDLALHAHETDHTTVDELVIHPTALHTHPSQS 982  
 QY 257 SIG-----GTRFVRYAEFDSR-CWLTQATV---RPPRAGLTVRVVTHGDSLVFLLT 305  
 DB 983 TVGAEFTDGHRIALHSRRTGTWVRHTTGLTSLNCTHPAALSTWSPT---DAHQIDLT 1039  
 QY 306 L-----SGPAFG 313  
 DB 1040 AYQQLADTGLAHYGPAPFG 1057

RESULT 3  
 A38096  
 perlecan precursor - human  
 N,Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate prote  
 C,Species: Homo sapiens (man)  
 C,Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text change 12-Jul-2004  
 A,Accession: A38096; S19256; S77946; A41059; A40306; A33625; A41736  
 R,Mutid: 267, 8544-8557, 1992  
 J, Biol. Chem. 267, 8544-8557, 1992  
 A,Title: Primary structure of the human heparan sulfate proteoglycan from basement membr  
 tor, laminin, neural cell adhesion molecules, and epidermal growth factor.  
 A,Reference number: A38096; MUID:92235084; PMID:1569102  
 A,Accession: A38096  
 A,Molecule type: mRNA  
 A,Residues: 1-4391 <MOR>  
 A,Cross-references: UNIPROT:P98160; GB:M85289; NID:G184426; PIDD:AAA52700.1; PID:G184427  
 J, Kallunki, P.; Trygvason, K.  
 J, Cell Biol. 116, 559-571, 1992  
 A,Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD pro  
 ell adhesion molecules, and epidermal growth factor.  
 A,Reference number: A41736; MUID:92112994; PMID:1730768  
 A,Accession: S19256  
 A,Molecule type: mRNA  
 A,Residues: 157, '59-434, 'A, '436, 'EU, '438-449, 'Q, '451-502, 'A, '503-792, 'K, '794-908, 'R, '7  
 71-2979, 'H, '2981-2994, 'G, '2996-3167, 'T, '3169-3240, 'R, '3242-3426, 'R, '3428-3631, 'Q, '3633-3  
 A,Cross-references: EMBL:X62515  
 R, Trygvason, K.

submitted to the EMBL Data Library, October 1991  
 A,Reference number: S77946  
 A,Accession: S77946  
 A,Molecule type: mRNA  
 A,Residues: 1-57, 'D, '59-434, 'A, '436, 'EU, '438-449, 'Q, '451-502, 'A, '503-792, 'K, '794-908, 'R, '7  
 71-2979, 'H, '2981-2994, 'G, '2996-3167, 'T, '3169-3240, 'R, '3242-3426, 'R, '3428-3631, 'Q, '3633-4  
 A,Cross-references: EMBL:X62515; NID:G29469; PIDD:CAA4373.1; PID:G29470  
 R,Kallunki, P.; Eddy, R.U.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Trygvason, K.  
 A,Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the ge  
 A,Reference number: A41059; MUID:92120660; PMID:1685141  
 A,Accession: A41059  
 A,Molecule type: mRNA  
 A,Residues: 'RT, '892-908, 'R, '910-1101, 'L, '1103-1132, 'L, '1134-1221, 'L, '1223-1397 <K2>  
 A,Cross-references: GB:S76436; NID:G243370; PIDD:AA821121.1; PID:G243371  
 R,Dodge, G.R.; Kovalesky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo, R  
 Genome 11, 673-680, 1991  
 A,Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellular  
 A,Reference number: A40306; MUID:91365376; PMID:1679745  
 A,Accession: A40306  
 A,Molecule type: mRNA  
 A,Residues: 1018-1405, 'G, '1407-1409, 'G, '1411-1465 <DOD>  
 A,Cross-references: GB:M64283; NID:G184424; PIDD:AAA5269.1; PID:G184425  
 R,Hermans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Caerman, J.J.; van den  
 J, Cell Biol. 109, 3199-3211, 1989  
 A,Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal  
 A,Accession: B33625  
 A,Reference number: A33625; MUID:90078352; PMID:2687294  
 A,Accession: B33625  
 A,Molecule type: protein  
 A,Residues: 1379-1384, 'X, '1386-1388, 'X, '1390-1398 <HE2>  
 A,Accession: A33625  
 A,Molecule type: protein  
 A,Residues: 2166-2171, 'X, '2173-2175, 'X, '2177-2185 <HE3>  
 A,Note: peptide potentially matches four different regions of sequence shown  
 C,Genetics: GDB:HSFG2  
 A,Map position: 1936.1-1936.1  
 C,Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembran  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-4391/Product: perlecan #status predicted <MAT>  
 F:22-193/Domain: I <DOM1>  
 F:194-530/Domain: II <DOM2>  
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
 F:531-1676/Domain: III <DOM3>  
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG>  
 F:1563-1610/Domain: laminin-type EGF-like homology <EG7>  
 F:1613-1668/Domain: laminin-type EGF-like homology <LEG8>  
 F:1677-3686/Domain: IV <DOM4>  
 F:2007-2034/Domain: transmembrane #status predicted <TM>  
 F:3687-4391/Domain: V <DOM5>  
 F:3845-3880/Domain: EGF homology <EGF1>  
 F:3888-3921/Domain: EGF homology <EGF>  
 F:3993-4106/Domain: laminin G repeat homology <LG2>  
 F:4147-4175/Domain: EGF homology <EGF2>  
 F:4149-4151/Region: motor neuron attachment (L-R-E) motif  
 F:4299-4301/Region: motor neuron attachment (L-R-E) motif  
 F:65, '71, '76/Binding site: heparan sulfate (Ser) (covalent) #status predicted (asn) (coval  
 F:89, '554, '1755, '2121, '3072, '3105, '3279, '3780, '3836, '4068/Binding site: carbohydrate (asn) (coval  
 F:2995, '3933, '4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 5.9%; Score 97; DB 2; Length 4391;  
 Best Local Similarity 22.3%; Pred. No. 31;  
 Matches 101; Conservative 42; Mismatches 133; Indels 176; Gaps 27;

QY 6 VLINSASGAN-----SIEGTA-----LPLPMLVHRTTRQDAFPVSWIKG----- 46  
 DB 2040 VVLSASDASPSPGVKIBSSPSTEGQTLNVCVAGSAHAQ-----VTVRRGSLPPT 2095

Qy 47 ---GDRSVAVLP-----HDPPFPAVHGDRDPL 74  
 Db 2096 QVHSSRLRLPQVSPADSGEYVCRVENGSGPKASITVSVLHGTSGSGSYTPFPSTR-PI 2154  
 Qy 75 LI-----AETL-----RQAMLVPHAGVGVGHFLMTLD- 105  
 Db 2155 RIBPSSSHVAGQTLDLNCVPGQAHQVTHKRGSGPARHQTHGSLRLHQVTPADSG 2214  
 Qy 106 -YTCGLDHLGVSG-EVAL------EVENACSQLKFRGQPV----- 139  
 Db 2215 EYVGHV-VGTSGLGLSVLVTIASVIRGPIPPVRISSSSYVAGQTLDLSCVVAQA 2272  
 Qy 140 QGVQVMAVRAAQLAA-----TGATTFE-TSP-----QVTRMRGDEATPTASVPGT- 186  
 Db 2273 HAQVTW-YKRGSGSLPARHQVGSRLYIFQASPADAGVYCRASNGMEASITVVTGTGA 2331  
 Qy 187 -APVPAARA-----GTRDEDVVLASSG-QVTMLR---VDSHPT----- 223  
 Db 2332 NLAVPAGSTOPRIEPPSSQVAGQTLDLNCVPGQAHQVTHKRGSLFVRHQTHGSL 2391  
 Qy 224 --LFQR-----PNDVPGMLLEAARQAACLV-T-GRAFPVPSIGTRFVR-----YAE- 268  
 Db 2392 LRLVQASPADSGEYVCRVENGSGPKASITVSVLHGTSGSGSYTPFPSTR-PI 2451  
 Qy 269 --FDSPCMIQATVVRPGPAAGLTVTVRVTGHQDG 298  
 Db 2452 QTLDLNCLV-----AGQAHQVTHKRG 2474

## RESULT 4

alpha-1-B-glycoprotein - human  
 C/Species: Homo sapiens (man)  
 C/Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 10-Sep-1999  
 C/Accession: A02113  
 R/Ichioka, N.; Takahashi, N.; Putnam, F.W.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 2363-2367, 1986  
 A/Title: Amino acid sequence of human plasma alpha-1B-glycoprotein: homology to the immu  
 A/Reference number: A02113; MUID:86205955; PMID:3458201  
 A/Accession: A02113  
 A/Molecule type: Protein  
 A/Residues: 1-474 <15>  
 C/Comment: The function of this plasma glycoprotein is unknown.  
 C/Genetics:  
 A/Gene: GDB:A1B3  
 A/Cross-references: GDB:119638; OMIM:138670  
 A/Map position: 19q-19q  
 C/Superfamily: alpha-1-B-glycoprotein; immunoglobulin homology  
 C/Keywords: duplication; glycoprotein; plasma  
 F/21-74/Domain: immunoglobulin homology <IM1>  
 F/11-163/Domain: immunoglobulin homology <IM2>  
 F/204-260/Domain: immunoglobulin homology <IM3>  
 F/297-355/Domain: immunoglobulin homology <IM4>  
 F/355-451/Domain: immunoglobulin homology <IM5>  
 F/23,159,342,350/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F/28-72,402-449/Disulfide bonds: #status predicted  
 F/118-161,211-258,304-353/Disulfide bonds: #status experimental

Query Match 5.9%; Score 96.5; DB 1; Length 474;  
 Best Local Similarity 23.3%; Pred. No. 2.3;  
 Matches 76; Conservative 37; Mismatches 142; Indels 71; Gaps 14;  
 Qy 8 INASDANSIEQTLALPVPMALVHRTVRQDAPVSGDRFSVTAVLPDHPFPAVH 67  
 Db 171 LSEPSATVTEELAPPPVLMHGB-----SSQVLLHPGNKVTTLTCAPLSGVDFOQR 224  
 Qy 68 GDRHDPFLIAETLRQAMLVPHAGVGVGHFLMTLDYTCGLDHLGVSGEVALEVEVA 127  
 Db 225 GEKE--LLVPRSSSPDRIFPHLN-AVALGSGHNTCYRLHNDNGMSGASAPVELLS 281  
 Qy 128 CS-----OLKFRGGQPVQGVDMAVRAAGRLAATGATTRTSQVTRRM 172  
 Db 282 DETLPAPERFSPESGRLRLRCLAPLEG-ARFALVREDR---GGRRVHRFQSPAGTEAL 337

Qy 173 RGFATPTASVPGTAPVPAARAGTRDEDVVLASSQDQTWRLRVDTSHPTLFGRPNDHV 232  
 Db 338 ---FELNISVADSANSYCVV-----DLKPPFGSGASPESELELHVDPPE---RPQLRA 386  
 Qy 233 PGMLLEAARQAACLVTPGAPVFP-----SIGTRFPRVYAFDSCWQATVR-PGPAAGL 287  
 Db 387 TWSGAVLAGDAVLRCEGPIPDVTFELLREGETKAVK-----TVRTPGAANL 434  
 Qy 288 TVRVVTGHQDGLVFLTLGSPAFSG 313  
 Db 435 -----ELIFV---GPOHAG 445

## RESULT 5

H75580  
 Conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
 C/Species: Deinococcus radiodurans  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Aug-2004  
 C/Accession: H75580  
 R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 266, 1571-1577, 1999  
 A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A/Reference number: A75250; MUID:20036896; PMID:10567266  
 A/Accession: H75580  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-323 <WHI>  
 A/Cross-references: UNIPROT:Q9RYN7; GB:AE001863; GB:AE001825; NID:96460670; PIDN:AAFL2484  
 A/Experimental source: strain R1  
 C/Genetics:  
 A/Gene: DRA0273  
 A/Map position: 2  
 C/Superfamily: mult domain homology

Query Match 5.7%; Score 94.5; DB 2; Length 323;  
 Best Local Similarity 24.3%; Pred. No. 2.1;  
 Matches 86; Conservative 24; Mismatches 121; Indels 123; Gaps 19;  
 Qy 2 PEAVVLINSASDANSIEQTLALPVPMALVHRTVRQDAPVSGDRFSVTAVLPDHP 61  
 Db 35 PRVLVLGSSANLARSVNNPFPAPREVEVFRALALDA-----GADLRREV- 78  
 Qy 62 FPAVHDDRDRPLIAETLRQAMLVPHAGVGVGHFLMTLDYTCGLDHLGVSGEVAE 121  
 Db 79 -FRPL-PDRPDADIMADVAVAEIF--GPGASVG-----LVG 113  
 Qy 122 LEVAVCSQK-FRGGQPVQGVDMAVRAAGRLAATGATTRPTS-----POVYRMR 173  
 Db 114 FENDASSSYLRWPFPGMRVN-----VPERGLNATIDIRALTBGALPAGLPAPVSGAL 167  
 Qy 174 GDEA-TFT-----ASVGTAPVPAARAGTRDEDVVLASSQDQTWRLR 216  
 Db 168 GRFAPHTPTCARLREMTAVGGERALPBGTV-----LQERN-LH 206  
 Qy 217 VDSHPTLFGRPND-----HVGCMLEAARQAACLVTPGAPVPSIGTRFVRVLEFD 270  
 Db 207 VMDRRVLAHRPGPIGGGLMELPGRVLPGE-----VLRPGEVLP-GEVPLPBGDTFD 259  
 Qy 271 SPCMIQAT-----VRPG-PAAGLTVRV-----TGHDGSLVFLTLGSG 308  
 Db 260 HPARTVTVRYVAQIRLGAAPRAGGVVPCLEVALSRRLHEDHS-VILARLTG 312

## RESULT 6

T36079  
 Hypothetical protein SCR126\_09 - Streptomyces coelicolor  
 C/Species: Streptomyces coelicolor  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C/Accession: T36079  
 R/Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A:Reference number: 221573  
A:Accession: U36079  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-362 <OLI>  
A:Cross-references: UNIPROT:O9X852; EMBL:AL049630; PIDD:CA040931.1; GSPDB:GN00070; SCOPED  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOE126.09

Query Match 5.7%; Score 93; DB 2; Length 362;  
Best Local Similarity 23.5%; Pred. No. 3.2; Indels 36; Gaps 9;  
Matches 52; Conservative 27; Mismatches 106; Indels 36; Gaps 9;

QY 110 LDHLGVSEVAVLEVEVACSQLKFRG-----GQPVQG-QVDWAVR-----RAGRLAAT 156  
DB 144 LDNVV--DVAVDEAGVAVRVLVVGDAERLGEVAVGAEVLAERWMPMAEGRRAEI 201  
QY 157 GTATTRFTSPQVYRRMRG----DEATPTASVGTAPVPAARAGTRD-----ED 201  
DB 202 GLPRDEAAVAAVADAGLAVADYAHVSARPPGTLTGFEGRTEPVPDGSCTITAH 261  
QY 202 VLSASSQODTWRLKVDTSHTPTLFORPNDHVPGLLEAARQAQCLV-TGPAPVPSIG- 259  
DB 262 VALDCAAAHTAKCTPEYAPNALTRPQREIRALGVGARPPALASTDPAAVVALAS 321  
QY 260 ---GTRFVRYAEFSDPCWIOATVRPQPAAGLTTVAVTGHD 297  
DB 332 ASQGAELTARGGGLGFRWLQPVGPVDAEAL-LVDVADHEE 361

## RESULT 7

202825  
Probable membrane protein L1231.3 [imported] - Leishmania major (strain Friedlin)  
C:Species: Leishmania major  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
C:Accession: E81455; 102825  
R:Myler, P.; Audleman, L.; devos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.;  
Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999  
A:Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c  
A:Reference number: A81455; MUID:99178987; PMID:10077609  
A:Accession: E81455  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-559 <PYL>  
A:Cross-references: UNIPROT:O00902; GB:AE001274; NID:g3264850; PIDD:AAC24648.1; PID:g226  
C:Genetics:  
A:Gene: L1231.3  
A:Map position: 1  
C:Superfamily: Leishmania major probable membrane protein L1231.3  
C:Keywords: transmembrane protein

Query Match 5.7%; Score 93; DB 2; Length 559;  
Best Local Similarity 21.2%; Pred. No. 5.4; Indels 115; Gaps 14;  
Matches 72; Conservative 39; Mismatches 115; Indels 114; Gaps 14;

QY 2 PEAVVLINSASDANSIEOTAL-----PVPMLVHTRVQDAFPVSWIPK 45  
DB 79 PPAVAAASVTSSSAAQSTTVAAAPGVPSGGAASSPVPSPVLETTQOQ-----PA 131  
QY 46 GGDRESVTAVALPHDHPFAP-----VHGDRHDPILLIETLRQAAML-VFHAQYGVV- 96  
DB 132 DDNRALAFAPARLIALVAPVPLVRLGLHRLHDMVSYDAAVHNASILRLQLAPYGIHAV 191  
QY 97 -----GHFLMTLDYTGHD-----HLGVSGEVALEVEVACSQLKFRGQGVQ 140  
DB 192 TGATRGCPFSLOAEYLCLTEAAPEVAVHPGAGGLAESD-----R 235  
QY 141 GOVDAAVVRAG-----RLAATGTATTRFTSPQVYR-----RMRGDPAT-----PTAS 182  
DB 236 GAVEADLINDHGDGASGRASAVSTAIAVTDPRTREARCSKRRARAAIQRDALMLRPPS 295

QY 183 VEPAPVPAARAGTRPDEVVLSA-----SSQODTWRLRV----- 217  
DB 296 LPSASSI-AAAASVSRBEDAASVAGGCGHRQAIRFCGLGSLRRRYCDRRSAFSSA 354  
QY 218 DTSHTPLFORPNDHVPGLLEAARQAQCLVTGPAPVPS 257  
DB 355 DVSPQVSYKVR-----RALESIVRCGYIVSAQAQPPPS 388

## RESULT 8

522155  
Oncogene 1 (tre-2 locus) (clone 210) - human  
N:Alternate names: transforming protein  
C:Species: Homo sapiens (man)  
C:Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 09-Jul-2004  
C:Accession: S57874; S22155  
R:Nakamura, T.; Hillova, J.; Mariage-Samson, R.; Onno, M.; Huebner, K.; Cannizzaro, L.A.;  
Oncogene 7, 733-741, 1992  
A:Title: A novel transcriptional unit of the tre oncogene widely expressed in human cancer  
A:Reference number: S57867; MUID:9228503; PMID:1565468  
A:Accession: S57874  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-786 <NAK>  
A:Cross-references: UNIPROT:Q15634; EMBL:X63546; NID:g37329; PIDD:CAA45108.1; PID:g37330

Query Match 5.7%; Score 93; DB 2; Length 786;  
Best Local Similarity 25.5%; Pred. No. 8.3;  
Matches 41; Conservative 21; Mismatches 61; Indels 38; Gaps 8;

QY 145 WAVRAGRLAATGATTRFTSPQVYRRMRGDFATPTASVGT---AVPAPARAGT----- 197  
DB 331 WAMNDTVLXKLRASTKLT-----RKQDLEPPARQGSGLAPRVVPAASRGKTLCKG 384  
QY 198 -----RDEPVLSASSQODTWRLKVDTSHTPTLFORPNDH-----VPGWLLEAA 241  
DB 385 YRQAPGPPAQFQRPICSA---PMAKSRSTPCPGAVDEDTYPVGTQGVPSIALAQGG 441  
QY 242 RQAA-CLVTGPAPVPS---IGTRFVRYAEFSDPCWIOA 277  
DB 442 PQGSMRFLEWMSMRPLPTDLDIGGPMFPHY-DFERSCWVBA 481

## RESULT 9

C75563  
Hypochemical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: C75563  
R:White, O.; Eissen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; May  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: C75563  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-307 <WHI>  
A:Cross-references: UNIPROT:Q9RYL8; GB:AE001863; GB:AE001825; NID:g6460670; PIDD:AAF1252  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRA0294  
A:Map position: 2  
C:Superfamily: Deinococcus radiodurans hypochemical protein DRA0294

Query Match 5.6%; Score 91.5; DB 2; Length 307;  
Best Local Similarity 27.2%; Pred. No. 3.5;  
Matches 61; Conservative 12; Mismatches 84; Indels 67; Gaps 11;

QY 113 LGVGEVALEVEVACSQLKFRGQPVQGVDM--AVRBAAGRLAA-TGTATTRFTSPQVY 169



Db 66 LGRNDVPAALTRAV-----LEERTGSSWCGPATAAQAALTLAARECTPALR-----QVT 117  
 Qy 170 RRMKGDPATPTASVPGTAPVAPABAGRTREDDVVLASASQOD-----TWRLRVDTS----- 220  
 Db 118 VRLGSPQOGLTVASAPTRLILIPAAKV-----QPGALISVQSSAPLAYAMELRVRRSGMNAQGS 172  
 Qy 221 -HPLFLQRPNDHVP-----GMLLEAPRQAACLVTPGAPFPVPSIGGTRFRVYAFD 270  
 Db 173 LSPFKIRQDKVTVMRDGIVTLTLKTPVRLTHLRVTDLF-----GGLEAVDDRPFD 227  
 Qy 271 SPCWIOAT-----VRPGPAAGLTVTR 291  
 Db 228 EPAMVNAAGSGDAVMADBSLYDRAVFLYEQVR-----GTTTTR 267

RESULT 10  
 G84309  
 hypothetical protein Vng1564h [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
 C:Accession: G84309  
 R:Ng, W.V.; Kennedy, S.P.; Mahitras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauer, B.; Keller, K.; Cruz, R.; Daneson, M.J.; Hough, D.W.; Maddocks, D.G.; Jadic Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; PMID:20504483; PMID:11016950  
 A:Accession: G84309  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-407 <STO>  
 A:Cross-references: UNIPROT:Q9HPM4; GB:AE004437; NID:g10581047; PIDN:AA019843.1; GSPDB:C C:Genetics:  
 A:Gene: VNG1564H

Query Match 5.6%; Score 91.5; DB 2; Length 407;  
 Best Local Similarity 22.0%; Pred. No. 4.9;  
 Matches 75; Conservative 22; Mismatches 97; Indels 147; Gaps 17;  
 Qy 42 WIRPGGRFSTYATLPHDHPFAPVHGDRDPLIAETLQAMLVTHAGVGVYHFL 101  
 Db 55 WRNAGDDRGSVVALAP-----FAG-GVVAAG----- 78  
 Qy 102 MTLDTYCHLDLHGVGEVLEVEVACSQLEKRGQPVQGVDAVPAVBAGLATGATTT 161  
 Db 79 -----ERGAAGEL-----RLHGE--TGAIVRRYAAAGDIGRP-OEDT 113  
 Qy 162 RFTSPQVYRRMGDPATPTASVPGTAPVAPABAGRTREDD-----VVLASASQODTWRL 215  
 Db 114 RFLPLPFV-----DVATQAGS--DRAFYAVRERYRGADGDAFQSAVYAFEGDNTPAMRF 166  
 Qy 216 RVDTSHTL-----FOR-PNDHVPQMLLEAPRQAACLVTPGAPFPVPSIGGTRFV 264  
 Db 167 RSDASPVSLADGDRVAVAFRCPCGDAAGVLVDAATG-----TQRRMDPTGGGDRRV 221  
 Qy 265 -----RYAEFDSPC-----WT-----QATVRGP-----AAGLTVV 290  
 Db 222 GDVALTRGVAAASHADYRGVCLDGDGVQVNVVIGRVERGEQTVVAYPMHVAHTGDSAV 281  
 Qy 291 RVTGHH-----ODGSLVFLTTLTG 308  
 Db 282 FVTGNTYRPDQRETDARHTENTHTAVGVGRGQVAFADVGG 322

RESULT 11  
 F70876  
 probable papA3 protein - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
 C:Accession: F70876  
 R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G. A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; PMID:9825987; PMID:9634220  
 A:Accession: F70876  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-472 <COL>  
 A:Cross-references: UNIPROT:O50438; GB:AL010186; GB:AL123456; NID:G3261493; PIDN:CAA1585; A:Experimental source: strain H37Rv  
 A:Genetics:  
 A:Gene: papA3

Query Match 5.6%; Score 91.5; DB 2; Length 472;  
 Best Local Similarity 22.0%; Pred. No. 5.9;  
 Matches 75; Conservative 29; Mismatches 122; Indels 115; Gaps 16;  
 Qy 6 VLINSASDANSIE-----QVALPVPMAVHRTVQDAFPVSWIRPKGDRFSTAVLPH 58  
 Db 111 IIRRTQDPADIFVFPVHGEITLPQIREIVQNT-----PDLQW-----GCFRFGIVQGDH 163  
 Qy 59 DHPFPAPVHGDRDPLIAETLQAMLVTHAGVGVYHFLMTL-----DYTGLDH 112  
 Db 164 -FTFFASVDHVDHVDAMVGVTLME-----FILMYAALVGGHAPLEPPAGSYDDFCRRQH 217  
 Qy 113 L-----GVSG-----EVAELE 123  
 Db 218 TFSSTLTVESPVQVANTKPAEGTNGSPDPPLPLGDPKSDADITYVMMLDSEQTHQFE 277  
 Qy 124 VEVASQLEKRGQPVQGVDAVPAVBAGLA--ATGATTRFTSPQVYRRMGDPATP- 179  
 Db 278 SVCTAGAGRTGG-----VLACCGLAHEHLGTTTYVGLTRDTRRTPADAMTGG 327  
 Qy 180 -TASVPGTAPV-----PAARAGRTREDDVVLASASQODTWRLRVDTSHTLFORPNDH 231  
 Db 328 WFGGLPIPIYPIPIGASFGDAARAQTSFDSGVLAIEVPYDR--VVELSSTLTMPRPN-- 382  
 Qy 232 VPGMLLEAPRQAACLVTPGAP--FVPSIGGTRFRVYAE 268  
 Db 383 FVYVNFUDA-----GAAPLSVLLTAEITGNTGIVYSD 414

RESULT 12  
 T26517  
 hypothetical protein Y18D10A.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C:Accession: T26517  
 R:Harris, B.  
 submitted to the EMBL Data Library, December 1998  
 A:Reference number: Z20226  
 A:Accession: T26517  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1634 <MIL>  
 A:Cross-references: UNIPROT:Q9XW25; EMBL:AL034393; PIDN:CAA22308.1; CESP:Y18D10A.1 A:Experimental source: clone Y18D10A  
 C:Genetics:  
 A:Gene: CESP:Y18D10A.1  
 A:Introns: 7/3; 44/3; 106/3; 193/1; 500/3; 533/3; 560/3; 590/3; 709/3; 1218/3; 1318/2; 14

Query Match 5.6%; Score 91.5; DB 2; Length 1634;  
 Best Local Similarity 28.4%; Pred. No. 27;  
 Matches 31; Conservative 12; Mismatches 39; Indels 27; Gaps 4;  
 Qy 171 RRMKGDPATPTASVPGTAPVAPABAGRTREDDVVLASAS-----SOODTWRLRVDTSHTLFOR 227  
 Db 426 RMAABASAGRSQAPGAPAPASASHLDPQDRLSMDPSGDSER-----EEL-----OK 475  
 Qy 228 PNDHVPQMLLEAPRQAACLV-----GAPFPVPSIGGTR 262  
 Db 476 QEGHIDAMVYTRRSARLSALPVPKKAASSSSKMPSPSPFSTRGRGR 524

## RESULT 13

B97566  
hypothetical protein AGR\_C\_3142 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C/Species: Agrobacterium tumefaciens  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C/Accession: B97566  
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A/Reference number: A97359; MUID:21608551; PMID:11743194  
A/Accession: B97566  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-912 <KUR>  
A/Cross-References: UNIPROT:O8UBP5; GB:AB007869; PIDD:AAK87483.1; PID:g15156806; GSPDB:C  
C/Genetics:  
A/Gene: AGR\_C\_3142  
A/Map position: circular chromosome

Query Match 5.5%; Score 91; DB 2; Length 912;

Best Local Similarity 20.2%; Pred. No. 15; Indels 94; Gaps 12;

Matches 60; Conservative 40; Mismatches 103; Indels 94; Gaps 12;

QY 25 PMALVHRTRVODAPFVSWIPKGGDRFSVTAVLPHDHPFAPVHGDRHDPDLIAETLRQA 84

DB 641 PVDVVGQRTILPEQLPL-----EGENDADMEAA-----GTPV-GETEDPRLISPEEKAA- 687

QY 85 MLVFNAGVGVPGVGHFLMTLDYTCCHDLGVSG-----EVALLEEVAVCSQLKFRGG 136

DB 688 -----HNEGAGATGVSPPKRVMTIVKPDGTLVAQEDVADPA 723

QY 137 QPVGQVDMVAVRAGRLAATGTATRTFTSPQVYRRMGDPA-----TPTASVPGT- 186

DB 724 QPPKAD-----KVALMLAPQTAKPGGAPVIAASRVPAEQAPQASTPAAKFSTQ 777

QY 187 --APVPAARAGRTDEDVVLSSQODTWRLRVDTSHPTLFQRPNDHVPGLLEAARQA 244

DB 778 SAAPLPAPARPS-SQPAVVAIVTQGNV-----RPAITTPA-----QAAQQQ 818

QY 245 ACLVTGPAPFVPSIGTRFVRYAEFDPSCWIOATVRGPAGLTVTVVTHGODSLV 301

DB 819 TAAATPAATSTPSAGG-----YYIQIASLPSQAELQSKSYQNMSAKFGSVI 863

## RESULT 14

AH2786  
conserved hypothetical protein Atu1710 [imported] - Agrobacterium tumefaciens (strain C5  
C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
C/Accession: AH2786  
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, I  
erage, G.; Gillet, W.; Grant, C.; Gunttner, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AH2786  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1008 <KUR>  
A/Cross-References: UNIPROT:O8UBP5; GB:AB008688; PIDD:AAI42710.1; PID:g17740147; GSPDB:C  
C/Genetics:  
A/Experimental source: strain C58 (Dupont)  
A/Gene: Atu1710  
A/Map position: circular chromosome

Query Match 5.5%; Score 91; DB 2; Length 1008;

Best Local Similarity 20.2%; Pred. No. 16;

Matches 60; Conservative 40; Mismatches 103; Indels 94; Gaps 12;

QY 25 PMALVHRTRVODAPFVSWIPKGGDRFSVTAVLPHDHPFAPVHGDRHDPDLIAETLRQA 84

DB 737 PVDVVGQRTILPEQLPL-----EGENDADMEAA-----GTPV-GETEDPRLISPEEKAA- 783

QY 85 MLVFNAGVGVPGVGHFLMTLDYTCCHDLGVSG-----EVALLEEVAVCSQLKFRGG 136

DB 784 -----HNEGAGATGVSPPKRVMTIVKPDGTLVAQEDVADPA 819

QY 137 QPVGQVDMVAVRAGRLAATGTATRTFTSPQVYRRMGDPA-----TPTASVPGT- 186

DB 820 QPPKAD-----KVALMLAPQTAKPGGAPVIAASRVPAEQAPQASTPAAKFSTQ 873

QY 187 --APVPAARAGRTDEDVVLSSQODTWRLRVDTSHPTLFQRPNDHVPGLLEAARQA 244

DB 874 SAAPLPAPARPS-SQPAVVAIVTQGNV-----RPAITTPA-----QAAQQQ 914

QY 245 ACLVTGPAPFVPSIGTRFVRYAEFDPSCWIOATVRGPAGLTVTVVTHGODSLV 301

DB 915 TAAATPAATSTPSAGG-----YYIQIASLPSQAELQSKSYQNMSAKFGSVI 959

## RESULT 15

AH0350  
Probable DNA-binding protein YP02880 [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis  
C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C/Accession: AH0350  
R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchell, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarragge, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; E  
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, E  
Nature 413, 523-527, 2001  
A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A/Reference number: AB0001; MUID:21470413; PMID:11586360  
A/Accession: AH0350  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-345 <KUR>  
A/Cross-References: UNIPROT:O8ZCT5; GB:AL590842; PIDD:CACT92131.1; PID:g15980847; GSPDB:G  
C/Genetics:  
A/Gene: YP02880

Query Match 5.5%; Score 90.5; DB 2; Length 345;

Best Local Similarity 24.3%; Pred. No. 4.9;

Matches 59; Conservative 27; Mismatches 90; Indels 67; Gaps 10;

QY 100 FLMTLDYTCCHDLGVSG-----EVALLEEVAVCSQLKFRGGQPV----- 139

DB 112 WLMSTWLVIVLVGLTGAMWQNHQAQAEIANK-VDQSAQLSQNGGQVPLTDDNSD 170

QY 140 -----QGVDMVAVRAGRLAATGTATT-----RFTSPQ-----YRRMGDFA 177

DB 171 AIAPTDAPAPVANOQPVPLTNHSGS-AITNSATTSVVKSTSTPVDATNTNTMHQGA 229

QY 178 TPTASVPGTAPVPAARAG-----TRDEDVVLSSAS-----QODTWRLRVDTSHPTLF 225

DB 230 ASAIVSFEQVPOPGWPTQPPPLPADAGVSGSASSVGLVNNFTADCMLOVVDAITGKTLF 289

QY 226 QRPNDHVPGLLEAARQAACLVTGPAPFVPSIGTRFVRYAEFDPSCWIOATVRGPAA 285

DB 290 ---SGIQKGAVLNLA-----GKAPYKLTIGAPGALITISYQGNPVLDISKFIKARVA 338

QY 286 GLT 288

DB 339 RLT 341

Search completed: July 20, 2005, 22:13:04

Job time : 44 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 21:51:49 ; Search time 45 Seconds  
(without alignments)  
519.226 Million cell updates/sec

Title: US-10-017-471B-17

Perfect score: 1645  
Sequence: 1 MPRAVVLINSASDANSIEQT.....GHODGLVFLTLSPAFSG 313

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	6.7	749	4	US-09-252-991A-20752
2	104	6.3	660	4	US-09-252-991A-32871
3	103	6.3	816	4	US-09-252-991A-23292
4	101	6.1	6396	4	US-09-410-551B-72
5	101	6.1	6396	4	US-09-940-316B-72
6	100	6.1	1257	4	US-09-252-991A-17290
7	99.5	6.0	632	4	US-09-252-991A-17148
8	97.5	5.9	638	4	US-09-252-991A-17917
9	97	5.9	4391	4	US-10-006-011A-2
10	94	5.7	239	4	US-09-252-991A-30746
11	94	5.7	676	4	US-09-948-016-7426
12	92.5	5.6	689	4	US-09-252-991A-31602
13	92.5	5.6	1490	4	US-09-252-991A-28442
14	92	5.6	480	4	US-09-252-991A-26799
15	92	5.6	794	4	US-09-252-991A-28569
16	91.5	5.5	565	4	US-09-252-991A-16757
17	91	5.5	374	4	US-09-252-991A-23565
18	90.5	5.5	424	4	US-09-252-991A-30209
19	90.5	5.5	548	4	US-10-096-571-12
20	90	5.5	738	4	US-09-252-991A-30721
21	89.5	5.4	466	4	US-09-684-855-113
22	89.5	5.4	466	4	US-09-684-855-136
23	89.5	5.4	466	4	US-09-684-855-159
24	89.5	5.4	466	4	US-09-488-265B-13
25	89.5	5.4	487	3	US-08-868-435-4
26	89.5	5.4	487	3	US-08-744-231-4
27	89.5	5.4	487	4	US-09-636-499-8

28	89.5	5.4	487	4	US-09-273-871A-15	Sequence 15, Appl
29	89.5	5.4	487	4	US-10-083-452-15	Sequence 15, Appl
30	89.5	5.4	487	4	US-09-635-504-14	Sequence 4, Appl
31	89.5	5.4	758	4	US-09-902-540-10176	Sequence 10176, A
32	89	5.4	353	4	US-09-252-991A-32769	Sequence 32769, A
33	89	5.4	398	4	US-09-252-991A-27517	Sequence 27517, A
34	89	5.4	2636	4	US-09-252-991A-25753	Sequence 25753, A
35	88	5.3	1841	2	US-08-804-227C-6	Sequence 6, Appl
36	88	5.3	4630	4	US-09-091-609-2	Sequence 2, Appl
37	88	5.3	5215	3	US-09-105-537-2	Sequence 2, Appl
38	87.5	5.3	316	4	US-09-902-540-10107	Sequence 10107, A
39	87.5	5.3	577	4	US-09-252-991A-19178	Sequence 19178, A
40	87	5.3	540	4	US-09-489-039A-7525	Sequence 7525, Ap
41	87	5.3	589	4	US-09-252-991A-32979	Sequence 32979, A
42	86.5	5.3	201	4	US-09-252-991A-18720	Sequence 18720, A
43	86.5	5.3	324	4	US-09-252-991A-24664	Sequence 24664, A
44	86	5.2	391	4	US-09-252-991A-20316	Sequence 20316, A
45	86	5.2	426	4	US-09-902-540-11847	Sequence 11847, A

#### ALIGNMENTS

```
RESULT 1
US-09-252-991A-20752
; Sequence 20752, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: MARC J. RUBENFIELD ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20752
; LENGTH: 749
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-20752

Query Match      6.7%; Score 110; DB 4; Length 749;
Best Local Similarity 24.7%; Pred. No. 0.011;
Matches 74; Conservative 26; Mismatches 116; Indels 84; Gaps 15;

QY 43 IPKGDRFSTAVLPHDHPFAPVGHDPDLIA-----ETLRQAMLVFHAGYGP 95
DB 465 VPERGDLAGMDVLGHGHLRDAGAGGHDALIVAGHQLLDLPGHFGANG-ARP 523
QY 96 VGYHFLMTLDYTCILDHIGVSGEVALEEV-EVACSQLEKFGQGVQGVQVMAV--RRAGR 152
DB 524 VGLH-----HHGLGERRVFPALPQLVAGQADAKQDHQVODDLAIQRGGX 571
QY 153 LAATGATTT-----RFTSPQVYRRMRGDFATPTASVVG--TAVPARARTR 198
DB 572 V--EGHATTWPPVSPGQWPGSRDAPALREARPPRPSPRPSGRLARAPFGP--RHR 626
QY 199 DEDVVLASASQDQWRRLRVDTSHPTLRFQRNDHVGMLLEAARQAACTVGP----- 251
DB 627 DRCAAPPAGAG-----SSCAA--RRPRRTGPKLSAPTRP--ITAPRRRRR 671
QY 252 ----APVPISIGTRFV----RYAEFDSF-----CWIQATVRGPAAGLTTR 291
DB 672 PARSGVPRPSRGRRRPPGRRRVIALDRPARARAGVARRSVCAVATVRTPPAGRCGNR 731

RESULT 2
US-09-252-991A-32871
; Sequence 32871, Application US/09252991A
```

Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32871  
; LENGTH: 660  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32871

Query Match 6.3%; Score 104; DB 4; Length 660;  
Best Local Similarity 25.9%; Pred. No. 0.041;  
Matches 72; Conservative 17; Mismatches 131; Indels 58; Gaps 12;

QY 34 VQDAPVWIPKGGRFSTAVLPDHPFPFAPVHGDRHDPILIAETLQAAMLVFH-AGY 92  
DB 1 IPSAPFPMRPPCCWPR-----RPSDDP---GVH-----PLQILPLRRPAAHLAHRPQM 46  
QY 93 GVPVGVHFLMTLDY-TCHLDHLGSGEVALEVEVACSLKFRGGQVPGQVDMVAV-BRA 150  
DB 47 SAARARHRLVLDGRRAGRLRSAGGALDLPQPPAREVHVRVPRRQAPGAGRL 106  
QY 151 GLAATGTAATRTFTSPQVYRRMRGDFATPTASVPGTAVPAPARAGRTDEDVLSASSQ 210  
DB 107 QODARTGCLLR---RRLFRGKGPLAQP-AVPPVGLPPLR----- 144  
QY 211 DTWRARVTSHTFLQRPNDHVPG-----MLLEAAQAALVYGPAPVPSIGGTFV 264  
DB 145 ---RLRTGSVPRGLDRDPDDIRGAALRQALRHPRGAFPSAAADQGSFGAGORR-- 199  
QY 265 RYAEFDSPCWIGATVTPG---PAAGLTVTVVTHODGS 299  
DB 200 -----GRACGIPRRVRPGGRRPAAAGADRRLQDQGPS 232

RESULT 3  
US-09-252-991A-23292  
; Sequence 23292, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23292  
; LENGTH: 816  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23292

Query Match 6.3%; Score 103; DB 4; Length 816;  
Best Local Similarity 31.0%; Pred. No. 0.071;  
Matches 58; Conservative 10; Mismatches 81; Indels 38; Gaps 10;  
QY 135 GGQPVQGVDMVVRAGR-LAATGTAATRTFTSPQVYR---RMRGDFATPTASVPGTAPVP 190  
DB 193 GAAAVHAGADAPDRRAGRLSLAAGSRTRPLHAQAARHLFGSRSG-----TDPG-APGD 244

QY 191 AARAGRTDEDEVLSASSQDPTWRLKRVDTSHPTLPFORPNHVPGMILLEAARQAALVTG 250  
DB 245 PLRQARRARRGQPAAAEGGSRSPAADR-----QRPARRPGRICRA----- 288  
QY 251 PAPVPSIGG---TRFVYAEFDSPCWIGATVPRGP-AAGLTVTVVTHODGSIVFTTL 306  
DB 289 PGRHPPLQAGGPRR-RTSRVDRP---GRLRGAAMARRRRVGRAGRRGGLCPGGPAP 342  
QY 307 SGPAFSG 313  
DB 343 SGPAVSG 349

RESULT 4  
US-09-410-551B-72  
; Sequence 72, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTU, DANIEL  
; APPLICANT: KU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; CURRENT FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 72  
; LENGTH: 6396  
; TYPE: PRT  
; ORGANISM: Streptomyces hygroscopicus  
US-09-410-551B-72

Query Match 6.1%; Score 101; DB 4; Length 6396;  
Best Local Similarity 26.3%; Pred. No. 2.5;  
Matches 97; Conservative 19; Mismatches 127; Indels 126; Gaps 22;

QY 19 QTALEVPMAALVHRRVQ---DAFPVSW-IPKGGRFSTV-AVLPHDHPFF-----APV 66  
DB 763 QITGPDEVRALHTALAOHLVHGVAILDWTLVIGSGRAPVTLPTYPFHDKDYLRLFTSRADV 822  
QY 67 HGDRHD-----PLLIAET-----LRQAAMLVFHA 90  
DB 823 TGAQGEQVYAHPLLGAAVAVLPGTGGVLTGRSLASHPWLGHAADVTLVLGAAFLIELAA 882  
QY 91 GYGVPPVGVHFLMTLDYTGCHLDH-----LGVSGEVALEVEVACSLKFRGGQVPG 141  
DB 883 RAGBEVG-----CDLHHELVETPPLVLPATGVVA-VSVHIAEPPDDTGRBAVTVHA 931  
QY 142 QVD-----WAVRAGRLAATGTAATRTFTSP-----QVYRRMRGDFATPTASV 183  
DB 932 RADSGGLM-TRHAGGFLGTAPAPATATDPAPMPAEAGPVADVDRFE-DIGSYSG-- 987  
QY 184 PGTAVPVPA--RAGRTDEDEVLSASSQDPTWRLKRVDTSHPTLPFORPNHVPGMILLEAAR 242  
DB 988 PGRGLRAAMRAGDTVYAEKALPDBQSADARF---TLHRA-----LIDDAF 1031  
QY 243 QAACTVGPAPFVPSIGGTRFVRVYAEFDSPCWIGATVPRGPAGLTVTVVTHODG--SL 300  
DB 1032 QAGLAALIDAF-----GGAARLPFSFQD-----VRI-HAAGATRLRLVTVGRDGERST 1077  
QY 301 VFLTTLSP 309

Db 1078 VRMT---GP 1083

## RESULT 5

US-09-940-316B-72  
 ; Sequence 72, Application US/09940316B  
 ; Patent No. 6759536  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KOSAN BIOSCIENCES, Inc.  
 ; APPLICANT: REEVES, CHRISTOPHER  
 ; APPLICANT: CHU, DANIEL  
 ; APPLICANT: KHOSLA, CHAITAN  
 ; APPLICANT: SANTI, DANIEL  
 ; APPLICANT: WU, KAI  
 ; TITLE OF INVENTION: POLYKETIDES ENCODING THE FKbA GENE OF THE FK-520 POLYKETIDE SYNTH  
 ; FILE REFERENCE: 30062-20026.11  
 ; CURRENT APPLICATION NUMBER: US/09/940,316B  
 ; PRIOR FILING DATE: 2001-08-27  
 ; PRIOR APPLICATION NUMBER: 09/410,551  
 ; PRIOR FILING DATE: 1999-10-01  
 ; PRIOR APPLICATION NUMBER: US 60/139,650  
 ; PRIOR FILING DATE: 1999-06-17  
 ; PRIOR APPLICATION NUMBER: US 60/123,810  
 ; PRIOR FILING DATE: 1999-03-11  
 ; PRIOR APPLICATION NUMBER: US 60/102,748  
 ; PRIOR FILING DATE: 1998-10-02  
 ; NUMBER OF SEQ ID NOS: 72  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 72  
 ; LENGTH: 6396  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces hygroscopicus  
 ; US-09-940-316B-72

Query Match 6.1%; Score 101; DB 4; Length 6396;

Best Local Similarity 26.3%; Pred. No. 2.5;  
 Matches 97; Conservative 19; Mismatches 127; Indels 126; Gaps 22;

QY 19 QTALEPMALVHRTVQ---DAFVSW-IRKGRFVSYT-AVLEPHDPFF-----APV 66  
 Db 763 QGTGPRDEVRALHTLAQILHVRGVAIDWTLVLVGGRRAPVLTPTYPFOHKDWLRPTSRADV 822  
 QY 67 HGRBHD-----PLIAET-----LROAMLVFHA 90  
 Db 823 TGAAGQEVANHPLLGAAVALPBTGSVVLTRLSLASHPMLGHAADVTVLLPGAAFLLEIAA 882  
 QY 91 GYGVVPVGHFLMTLDYTCNLDH-----LGVSGEVALEVEVACSQLKFRGGQPVQG 141  
 Db 883 RAGDEVG-----CDLHELVITETPLVLPATGVAA-VSVEIAEPDDTGRRAVYVHA 931  
 QY 142 QVD---WAVRAGRLAATGATATFTSP-----OYRRMGDFATPTASV 183  
 Db 932 RADSSGLM-TRHAGGFLGTAPAPATATDPAPWPAEAGPVDAVDYDFE-DIGSYG-- 987  
 QY 184 PGTAPVPA--RAGTRDEVDVLSASSQODTWRLAVDTSHTPLFRPNDHPGMLLEAAR 242  
 Db 988 PGRFGLAAMRAGDTVAEVALPDEGADAAFP--TLHP--LDDAAF 1031  
 QY 243 QAACTLVGAPAPVSGITGRFVRAEFPSPCMIOATVPAPAGLTTVTVTGHDG--SL 300  
 Db 1032 QAGGLAALDAP-----GGAARLPSPFD-----VRI-HAAGATRLRTVVGDSGERST 1077  
 QY 301 VFLTTLSP 309  
 Db 1078 VRMT---GP 1083

RESULT 6  
 US-09-252-991A-17290  
 ; Sequence 17290, Application US/09252991A  
 ; Patent No. 6551795

GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196,136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 17290  
 ; LENGTH: 1257  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-17290

Query Match 6.1%; Score 100; DB 4; Length 1257;  
 Best Local Similarity 23.3%; Pred. No. 0.28;  
 Matches 73; Conservative 26; Mismatches 114; Indels 100; Gaps 13;

QY 3 EAVVLNSASDANSIQTALPVPMALVHRT-RVQDAFPVSMIRKGD--RFSVTAVLP-- 57  
 Db 806 DRTAFISAGSVQAQQRAFLVGVLEHVAEELAEF-----AQGGELLEAVDQVAPRR 860  
 QY 58 -----HDHPFF---APVHGRHDPPLIAETLRQAMLVFH 89  
 Db 861 VALLQAFHQARSETLEPGEFAVEHRQRAFRALAVVEGSHPRDLRAFLAVQAVEELHE 920  
 QY 90 AGYGVVPVGH-----FLMTLDYTCNLDHLSVSGEVA-----ELEVEVA 127  
 Db 921 ARHQVELGQHVDHREARPELVLQDLAC-ADRLGMGALGVAAQQVKGADGERALDGA 979  
 QY 128 CSQKFRGGQPVQ--GVYMAVRAGRLAATGATATFTSPQYRRKRGDPATPTASVP 165  
 Db 980 ALALGLEQVEAEAPPGGLVHGVAVLGGIAAGVQOHRFV-----G 1019  
 QY 186 TAPVPAARAGRTDE-----DVLSASSQODTWRLAVDTSHTPLFRPNDHPGMLL 238  
 Db 1020 EPPVAVARAADPDPRCAHLGGEREVQAGVHQGRLARTG-----RADDHVPQQLVQ 1071  
 QY 239 EAARQAACLVTPG 251  
 Db 1072 VAPLAAAGL--GP 1082

## RESULT 7

US-09-252-991A-17148  
 ; Sequence 17148, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196,136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 17148  
 ; LENGTH: 632  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-17148

Query Match 6.0%; Score 99.5; DB 4; Length 632;  
 Best Local Similarity 23.7%; Pred. No. 0.11;  
 Matches 66; Conservative 26; Mismatches 108; Indels 79; Gaps 12;





```

US-09-252-991A-26799
; Sequence 26799, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26799
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26799

```

```

Query Match      5.6%; Score 92; DB 4; Length 480;
Best Local Similarity 28.1%; Pred. No. 0.48;
Matches 56; Conservative 17; Mismatches 76; Indels 50; Gaps 12;

```

```

QY      129 SQLKRGQGVQGVDMVAVRAGRL---AATGATRTFTSPQY-----RMRGDF 176
      204 ARLGRGARPARRRPARFVLRQQLPAGHPGTQAGHRDQVLGCGCAGPGRQSGQ 263
DB      177 ATEPTASV-----PGR---APVPAARAGRTRDEDVL-----SASSQDTWRLRYD---T 219
      264 RQPGAPVGRADPGABRSPPV--PGRHPELRIRQCLRSATSRHSGHLRADPALR 321
QY      220 SHPTLFORPNDHVPGLLEAARQA---ACTVGPAPFVPSIGTRFYVAEFDSPC--W 274
      322 AHPVHFVQPGN--PGAENHSTARHAPDQPLRAVRVPLHPHGRKRRPAR-----TPLEW 374
DB      275 IQATVRP-----GPAG 286
      375 PRPTARPCLRVAPGGAAPG 393

```

```

RESULT 15
US-09-252-991A-28569
; Sequence 28569, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28569
; LENGTH: 794
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28569

```

```

Query Match      5.6%; Score 92; DB 4; Length 794;
Best Local Similarity 25.7%; Pred. No. 1;
Matches 56; Conservative 11; Mismatches 77; Indels 74; Gaps 12;

```

```

QY      135 GQPVQGVQVDMVAVRAGRLAATGATRTFTSPQVYRRMRGDFATPTASVGTAPVP-AAR 193
      102 GSRPISAPA-ISRMRGSRACGSGTSTP--TAPATAR-----AASASTPWPAPWPMATN 153
QY      194 AGRTRD--EDVVISASSQDTWRLRYDTS-----HPTLFORPND---HVPG--MLLEA 240

```

```

DB      154 RGRSASFPTSPVSMASRSRTTWBATGAPAPRPMWSPTRMATPTSTDRASVPGNYAMTWTS 213
      241 AROA---ACT-----VTGPAPV-----
DB      214 ARSACLSACRCWCPAIVPGAPAVPPRASTPDSTGVGTGKTSAPISGRAPSPAS 273
QY      256 -----PSIGTRFVRYAEFDSPCWQAT--VRPGPAG 286
      274 RCAPRRPGTGATRAISTATSTKPGWSSITPARSGDAPG 311

```

```

Search completed: July 20, 2005, 22:13:57
Job time : 47 secs

```









```
RESULT 3
US-10-017-471A-19
; Sequence 19, Application US/10017471A
; Publication No. US20030124644A1
; GENERAL INFORMATION:
; APPLICANT: Takano, Eriko
; APPLICANT: Bidd, Mervyn
; TITLE OF INVENTION: Antibiotic Production
; FILE REFERENCE: 0380-P02329US1
; CURRENT APPLICATION NUMBER: US/10/017,471A
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,561
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 4346
; TYPE: DNA
; ORGANISM: Streptomyces coelicolor
US-10-017-471A-19

Query Match          80.0%; Score 19.2; DB 15; Length 4346;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1  GTCTGTGTCGCGGTGACGCGSAC 24
         |||||:||||:||||:||||:
DB      1250 GTCTGTGATGCCGCGTACCGCGCAC 1273

RESULT 4
US-10-369-493-39801/c
; Sequence 39801, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 39801
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-10-369-493-39801

Query Match          78.3%; Score 18.8; DB 17; Length 1100;
Best Local Similarity 70.8%; Pred. No. 83;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1  GTCTGTGTCGCGGTGACGCGSAC 24
         |||||:||||:||||:||||:
DB      603  GTCTGTGTTGCCGCGTACCGACGAC 580

RESULT 5
US-10-369-493-39049/c
; Sequence 39049, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
```

```
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 39049
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-10-369-493-39049

Query Match          78.3%; Score 18.8; DB 17; Length 1136;
Best Local Similarity 70.8%; Pred. No. 82;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1  GTCTGTGTCGCGGTGACGCGSAC 24
         |||||:||||:||||:||||:
DB      639  GTCTGTGTTGCCGCGTACCGACGAC 616

RESULT 6
US-10-369-493-39416/c
; Sequence 39416, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 39416
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-10-369-493-39416

Query Match          78.3%; Score 18.8; DB 17; Length 1136;
Best Local Similarity 70.8%; Pred. No. 82;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1  GTCTGTGTCGCGGTGACGCGSAC 24
         |||||:||||:||||:||||:
DB      639  GTCTGTGTTGCCGCGTACCGACGAC 616

RESULT 7
US-10-437-963-98196
; Sequence 98196, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 98196  
LENGTH: 2508  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_96124C.1  
US-10-437-963-98196

Query Match 78.3%; Score 18.8; DB 19; Length 2508;  
Best Local Similarity 70.8%; Pred. No. 70;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GTCCGTGCGCCGCGTACGCGSAC 24  
456 GTCCGTGCGCCGCGTACGCGSAC 479

## RESULT 8

US-10-282-122A-30494/C  
Sequence 30494, Application US/10282122A  
Publication No. US20040029129A1

## GENERAL INFORMATION:

APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zykund, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: EUTRA.034A

CURRENT APPLICATION NUMBER: US/0/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 30494

LENGTH: 4254

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-10-282-122A-30494

Query Match 78.3%; Score 18.8; DB 17; Length 4254;  
Best Local Similarity 70.8%; Pred. No. 63;

Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GTCCGTGCGCCGCGTACGCGSAC 24  
927 GTCCGTGCGCCGCGTACGCGSAC 904

## RESULT 9

US-10-424-599-23810/C  
Sequence 23810, Application US/10424599  
Publication No. US20040031072A1

## GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 23810  
LENGTH: 365  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_121501C.1  
US-10-424-599-23810

Query Match 74.2%; Score 17.8; DB 18; Length 365;  
Best Local Similarity 69.6%; Pred. No. 2,8e+02;  
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GTCCGTGCGCCGCGTACGCGSAC 23  
260 GTCCGTGCGCCGCGTACGCGSAC 238

## RESULT 10

US-10-425-115-70440  
Sequence 70440, Application US/10425115  
Publication No. US20040214272A1

## GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 70440  
LENGTH: 560  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(560)  
OTHER INFORMATION: unsure at all n locations  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_164230C.1  
US-10-425-115-70440

Query Match 74.2%; Score 17.8; DB 20; Length 560;  
Best Local Similarity 69.6%; Pred. No. 2,5e+02;  
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GTCCGTGCGCCGCGTACGCGSAC 23  
213 GTCCGTGCGCCGCGTACGCGSAC 235

```
RESULT 11
US-10-956-157-187
; Sequence 187, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 187
; LENGTH: 737
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-187

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 21; Length 737;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTGTGGCCGCGTACGCGGAC 24
DB 678 TCCTGTGGCCGCGTACGCGGAC 700

RESULT 12
US-10-956-157-5422/c
; Sequence 5422, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5422
; LENGTH: 737
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-5422

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 21; Length 737;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTGTGGCCGCGTACGCGGAC 24
DB 60 TCCTGTGGCCGCGTACGCGGAC 38

RESULT 13
US-10-027-632-156446
; Sequence 156446, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; NUMBER OF SEQ ID NOS: 2002-04-30
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 156446
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156446

Query Match
Best Local Similarity 76.2%; Score 17.8; DB 13; Length 962;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156446
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156446

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 13; Length 962;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 CTGTGGCCGCGTACGCGGAC 24
DB 544 CTGTGGCCGCGTACGCGGAC 564

RESULT 14
US-10-027-632-156447
; Sequence 156447, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; NUMBER OF SEQ ID NOS: 2002-04-30
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 156447
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156447

Query Match
Best Local Similarity 76.2%; Score 17.8; DB 13; Length 962;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 CTGTGGCCGCGTACGCGGAC 24
DB 544 CTGTGGCCGCGTACGCGGAC 564

RESULT 15
US-10-027-632-156448
```

```
Sequence 156448, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156448
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156448

Query Match          74.2%; Score 17.8; DB 13; Length 962;
Best Local Similarity 76.2%; Pred. No. 2.3e+02;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      4 CTGCTGCGCGGTACGCGCAC 24
Db      544 CTGCTGCGCGGTACGCGCAC 564
```

```
RESULT 16
US-10-027-632-156446
; Sequence 156446, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156446
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156446
```

```
Query Match          74.2%; Score 17.8; DB 17; Length 962;
Best Local Similarity 76.2%; Pred. No. 2.3e+02;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      4 CTGCTGCGCGGTACGCGCAC 24
Db      544 CTGCTGCGCGGTACGCGCAC 564
```

```
RESULT 17
US-10-027-632-156447
; Sequence 156447, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156447
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156447
```

```
Query Match          74.2%; Score 17.8; DB 17; Length 962;
Best Local Similarity 76.2%; Pred. No. 2.3e+02;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      4 CTGCTGCGCGGTACGCGCAC 24
Db      544 CTGCTGCGCGGTACGCGCAC 564
```

```
RESULT 18
US-10-027-632-156448
; Sequence 156448, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
```

```

; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156448
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156448

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 17; Length 962;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 GTGCTGCGCGGTACCGCGAC 24
DB 544 GTGCTGCGCGGTACCGCGAC 564

RESULT 19
US-10-156-761-1939
; Sequence 1939, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1939
; LENGTH: 1557
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1557)
US-10-156-761-1939

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 15; Length 1557;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGCGCGGTACCGCGA 23
DB 190 GTCTGTGCGCGGTACCGCGA 212

RESULT 20
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
```

```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 15; Length 9025608;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGCGCGGTACCGCGA 23
DB 2388161 GTCTGTGCGCGGTACCGCGA 2388139

RESULT 21
US-10-027-632-127914/c
; Sequence 127914, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127914
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-127914

Query Match
Best Local Similarity 73.3%; Score 17.6; DB 13; Length 413;
Matches 17; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTCTGTGCGCGGTACCGCGA 24
DB 281 GTCTGTGCGCGGTACCGCGA 258

RESULT 22
US-10-027-632-127914/c
; Sequence 127914, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```



```
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-28
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 127914
LENGTH: 413
TYPE: DNA
ORGANISM: Human
US-10-027-632-127914
```

```
Query Match 73.3% Score 17.6; DB 17; Length 413;
Best Local Similarity 70.8% Pred. No. 3.3e+02;
Matches 17; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
Oy 1 GTCTGTGCGCGCGTSGACGCGSAC 24
Db 281 GTCTGTGCGCGCGTSGACGCGCAC 258
```

```
RESULT 23
US-10-437-963-45501
Sequence 45501, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 45501
LENGTH: 4086
TYPE: DNA
ORGANISM: Oryza sativa
FEATURES:
OTHER INFORMATION: Clone ID: PAT_MRT4530_48460C.1
US-10-437-963-45501
```

```
Query Match 73.3% Score 17.6; DB 19; Length 4086;
Best Local Similarity 70.8% Pred. No. 2.1e+02;
Matches 17; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
Oy 1 GTCTGTGCGCGCGTSGACGCGSAC 24
Db 141 GTCTGTGCGCGCGCGTSGACGCGCAC 164
```

```
RESULT 24
US-09-815-242-7762
Sequence 7762, Application US/09815242
Patent No. US2002006159A1
GENERAL INFORMATION:
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7762
LENGTH: 744
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURES:
NAME/KEY: CDS
LOCATION: (1)...(744)
US-09-815-242-7762
```

```
Query Match 72.5% Score 17.4; DB 9; Length 744;
Best Local Similarity 71.4% Pred. No. 3.6e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 2 TCCTGTGCGCGCGTSGACGCGS 22
Db 602 TCCTGTGCGCGCGTSGACGCGG 622
```

```
RESULT 25
US-10-282-122A-30168
Sequence 30168, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
```

```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 30168
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30168

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 17; Length 744;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCCTGTGGCCGCGTSAACGCGS 22
    |||||:|||||:|||||:|||||:
Db 602 TCCTGTGGCCGCGTSAACGCGG 622

RESULT 26
US-10-739-930-2160
; Sequence 2160, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: KOVALIC, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 2160
; LENGTH: 1042
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-CLUSTER281_2
US-10-739-930-2160

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 20; Length 1042;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCTGTGGCCGCGTSAACGCGSA 23
    |||||:|||||:|||||:|||||:
Db 375 CCTGTGGCCGCGTSAACGCGCA 395

RESULT 27
US-10-425-115-60583
; Sequence 60583, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 60583
; LENGTH: 1348
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_155251C.1
US-10-425-115-60583

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 20; Length 1348;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCTGTGGCCGCGTSAACGCGSA 23
    |||||:|||||:|||||:|||||:
Db 479 CCTGTGGCCGCGTSAACGCGCA 499

RESULT 28
US-10-156-761-2779
; Sequence 2779, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2779
; LENGTH: 1719
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1719)
US-10-156-761-2779

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 15; Length 1719;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CTGTGGCCGCGTSAACGCGSAC 24
    |||||:|||||:|||||:|||||:
Db 1681 CTGTGGCCGCGTSAACGCGGAC 1701

RESULT 29
US-09-815-242-7690/C
; Sequence 7690, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Kari L.
; APPLICANT: Zykkind, Judith W.
```

APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7690  
LENGTH: 2553  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(2553)  
US-09-815-242-7690

Query Match 72.5%; Score 17.4; DB 9; Length 2553;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTGCTGCCGCGATCACC 22  
DB 1550 TCCTGCTGCCGCGATCACC 1530

RESULT 30  
US-10-282-122A-29994/C  
Sequence 29994, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zykkind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 29994  
LENGTH: 2553  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-10-282-122A-29994

Query Match 72.5%; Score 17.4; DB 17; Length 2553;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTGCTGCCGCGATCACC 22  
DB 1550 TCCTGCTGCCGCGATCACC 1530

Search completed: July 20, 2005, 22:08:55  
Job time : 605.143 secs

**THIS PAGE IS BLANK**



101	16.2	67.5	1624	3	US-08-582-740-67	Sequence 67, Appl	174	16	66.7	12253	4	US-09-902-540-1060	Sequence 1060, Ap
102	16.2	67.5	1624	3	US-09-109-879-67	Sequence 67, Appl	175	15.8	65.8	224	4	US-09-513-999C-20483	Sequence 20483, A
103	16.2	67.5	1707	4	US-09-266-965-68	Sequence 68, Appl	176	15.8	65.8	330	4	US-09-489-039A-5929	Sequence 5929, Ap
104	16.2	67.5	1809	1	US-08-453-956-24	Sequence 24, Appl	177	15.8	65.8	420	4	US-09-902-540-2686	Sequence 2686, Ap
105	16.2	67.5	1809	1	US-08-086-631-24	Sequence 24, Appl	178	15.8	65.8	546	4	US-09-902-540-1118	Sequence 7118, Ap
106	16.2	67.5	1809	1	US-08-452-930-24	Sequence 24, Appl	179	15.8	65.8	588	4	US-09-252-991A-7779	Sequence 7779, Ap
107	16.2	67.5	1809	5	PCT-US93-08174-24	Sequence 24, Appl	180	15.8	65.8	601	4	US-09-949-016-131627	Sequence 131627, A
108	16.2	67.5	2064	4	US-09-489-039A-5042	Sequence 5042, Ap	181	15.8	65.8	640	4	US-09-902-540-65	Sequence 65, Appl
109	16.2	67.5	2136	4	US-09-489-039A-5017	Sequence 5017, Ap	182	15.8	65.8	645	4	US-09-489-039A-5702	Sequence 5702, Ap
110	16.2	67.5	2178	4	US-09-902-540-6462	Sequence 6462, Ap	183	15.8	65.8	768	4	US-09-252-991A-9790	Sequence 9790, Ap
111	16.2	67.5	2387	4	US-09-902-540-453	Sequence 453, App	184	15.8	65.8	801	4	US-09-252-991A-10049	Sequence 10049, A
112	16.2	67.5	2421	4	US-09-328-352-3034	Sequence 3034, Ap	185	15.8	65.8	804	4	US-09-252-991A-7631	Sequence 7631, Ap
113	16.2	67.5	2972	2	US-08-720-484A-3	Sequence 3, Appl	186	15.8	65.8	1049	3	US-09-072-967-303	Sequence 303, App
114	16.2	67.5	2972	3	US-08-953-823A-3	Sequence 3, Appl	187	15.8	65.8	1049	4	US-09-902-540-2952	Sequence 2952, Ap
115	16.2	67.5	2972	3	US-09-398-239-3	Sequence 3, Appl	188	15.8	65.8	1119	4	US-09-902-540-1119	Sequence 2952, Ap
116	16.2	67.5	2972	4	US-09-560-876A-3	Sequence 3, Appl	189	15.8	65.8	1155	2	US-08-387-942C-20	Sequence 20, Appl
117	16.2	67.5	3033	4	US-09-724-797-81	Sequence 81, Appl	190	15.8	65.8	1162	4	US-09-016-444-1240	Sequence 1240, Ap
118	16.2	67.5	3402	4	US-09-252-991A-15560	Sequence 15560, A	191	15.8	65.8	1188	4	US-09-489-039A-1535	Sequence 1535, Ap
119	16.2	67.5	3606	4	US-09-252-991A-15688	Sequence 15688, A	192	15.8	65.8	1269	4	US-09-894-844-115	Sequence 115, App
120	16.2	67.5	3921	4	US-09-949-016-628	Sequence 628, App	193	15.8	65.8	1335	4	US-09-252-991A-765	Sequence 765, App
121	16.2	67.5	3921	4	US-09-949-016-2120	Sequence 2120, Ap	194	15.8	65.8	1338	4	US-09-902-540-1085	Sequence 9790, Ap
122	16.2	67.5	4032	1	US-08-126-587C-8	Sequence 8, Appl	195	15.8	65.8	1479	4	US-09-489-039A-1852	Sequence 1852, Ap
123	16.2	67.5	4194	4	US-09-902-540-2099	Sequence 2099, App	196	15.8	65.8	1485	4	US-09-489-039A-2769	Sequence 2769, Ap
124	16.2	67.5	4195	4	US-09-902-540-635	Sequence 635, App	197	15.8	65.8	1524	4	US-09-252-991A-737	Sequence 737, App
125	16.2	67.5	4824	4	US-09-949-016-1068	Sequence 1068, App	198	15.8	65.8	1627	4	US-09-022-655-515	Sequence 515, App
126	16.2	67.5	4902	4	US-09-902-540-5535	Sequence 5635, Ap	199	15.8	65.8	1641	4	US-09-252-991A-646	Sequence 646, App
127	16.2	67.5	5559	1	US-08-287-442-3	Sequence 3, Appl	200	15.8	65.8	1761	4	US-09-902-540-7394	Sequence 7394, Ap
128	16.2	67.5	5559	1	US-08-459-701-3	Sequence 3, Appl	201	15.8	65.8	1887	4	US-09-252-991A-12065	Sequence 12065, A
129	16.2	67.5	5559	1	US-08-460-298-3	Sequence 3, Appl	202	15.8	65.8	2091	4	US-09-252-991A-12065	Sequence 12065, A
130	16.2	67.5	5559	1	US-08-459-174-3	Sequence 3, Appl	203	15.8	65.8	2106	4	US-09-902-540-6818	Sequence 6818, Ap
131	16.2	67.5	5559	5	PCT-US93-06300A-3	Sequence 3, Appl	204	15.8	65.8	2142	4	US-09-902-540-9193	Sequence 9193, Ap
132	16.2	67.5	5672	4	US-09-023-655-1392	Sequence 1392, Ap	205	15.8	65.8	2181	4	US-09-252-991A-11948	Sequence 11948, A
133	16.2	67.5	7943	4	US-09-949-016-15735	Sequence 15735, A	206	15.8	65.8	2583	4	US-09-252-991A-12036	Sequence 12036, A
134	16.2	67.5	9097	4	US-09-902-540-947	Sequence 947, App	207	15.8	65.8	2634	3	US-08-941-936-1	Sequence 1, Appl
135	16.2	67.5	10763	1	US-08-761-258-1	Sequence 1, Appl	208	15.8	65.8	4048	4	US-09-902-540-6567	Sequence 567, App
136	16.2	67.5	10763	2	US-08-977-306-1	Sequence 1, Appl	209	15.8	65.8	4412	4	US-09-949-016-6565	Sequence 5656, App
137	16.2	67.5	11219	1	US-07-642-734C-1	Sequence 1, Appl	210	15.8	65.8	4689	3	US-09-105-337-34	Sequence 34, Appl
138	16.2	67.5	11219	3	US-08-439-009A-1	Sequence 1, Appl	211	15.8	65.8	4714	4	US-09-949-016-1137	Sequence 1137, Ap
139	16.2	67.5	12419	4	US-09-902-540-1059	Sequence 1059, App	212	15.8	65.8	5234	4	US-09-902-540-710	Sequence 710, App
140	16.2	67.5	12840	4	US-09-949-016-15720	Sequence 15720, A	213	15.8	65.8	9039	4	US-09-902-540-983	Sequence 983, App
141	16.2	67.5	12840	4	US-09-949-016-15720	Sequence 12817, A	214	15.8	65.8	11029	4	US-09-902-540-1004	Sequence 1004, Ap
142	16.2	67.5	17476	4	US-09-949-016-16865	Sequence 16865, A	215	15.8	65.8	12588	2	US-08-387-942C-1	Sequence 1, Appl
143	16.2	67.5	17622	4	US-09-902-540-1148	Sequence 1128, App	216	15.8	65.8	12665	4	US-09-902-540-1126	Sequence 1126, App
144	16.2	67.5	17622	4	US-09-902-540-1148	Sequence 1148, App	217	15.8	65.8	12665	4	US-09-949-016-12309	Sequence 12309, Ap
145	16.2	67.5	18031	4	US-09-902-540-1180	Sequence 1180, App	218	15.8	65.8	19980	4	US-09-949-016-13533	Sequence 13533, Ap
146	16.2	67.5	18318	1	US-08-414-926A-6	Sequence 6, Appl	219	15.8	65.8	25668	4	US-09-902-540-1246	Sequence 1246, Ap
147	16.2	67.5	18318	2	US-08-926-922-6	Sequence 6, Appl	220	15.8	65.8	25992	4	US-09-949-016-17425	Sequence 17425, A
148	16.2	67.5	18318	3	US-09-253-682-6	Sequence 6, Appl	221	15.8	65.8	26194	3	US-09-902-540-1250	Sequence 1250, App
149	16.2	67.5	18318	3	US-09-527-657-6	Sequence 6, Appl	222	15.8	65.8	32198	3	US-09-105-537-5	Sequence 5, Appl
150	16.2	67.5	18318	4	US-09-892-100-6	Sequence 6, Appl	223	15.8	65.8	35506	3	US-09-320-878-19	Sequence 19, Appl
151	16.2	67.5	24535	4	US-09-949-016-12810	Sequence 12810, A	224	15.8	65.8	38506	4	US-09-141-908-1	Sequence 1, Appl
152	16.2	67.5	25733	4	US-09-902-540-1215	Sequence 1215, App	225	15.8	65.8	38506	4	US-09-657-440-19	Sequence 19, Appl
153	16.2	67.5	31039	4	US-09-949-016-13449	Sequence 13449, A	226	15.8	65.8	38961	4	US-09-949-016-12143	Sequence 12143, A
154	16.2	67.5	36203	4	US-09-949-016-15430	Sequence 15430, A	227	15.8	65.8	38961	4	US-09-949-016-14474	Sequence 14474, A
155	16.2	67.5	36387	4	US-09-949-016-12370	Sequence 12370, A	228	15.8	65.8	33243	4	US-09-949-016-15443	Sequence 15443, A
156	16.2	67.5	36387	4	US-09-949-016-13862	Sequence 13862, A	229	15.8	65.8	38243	4	US-09-949-016-15937	Sequence 15937, A
157	16.2	67.5	47981	4	US-09-679-279-1	Sequence 1, Appl	230	15.8	65.8	82494	4	US-09-949-016-16302	Sequence 16302, A
158	16.2	67.5	49255	4	US-09-902-540-1269	Sequence 1269, App	231	15.8	65.8	88758	4	US-09-949-016-17398	Sequence 17398, A
159	16.2	67.5	53500	4	US-09-266-965-76	Sequence 76, Appl	232	15.8	65.8	133159	4	US-09-949-016-12879	Sequence 12879, A
160	16.2	67.5	4403765	3	US-09-103-840A-2	Sequence 2, Appl	233	15.8	65.8	133577	4	US-09-949-016-12879	Sequence 12879, A
161	16.2	67.5	441529	3	US-09-103-840A-1	Sequence 1, Appl	234	15.8	65.8	133577	4	US-09-949-016-12879	Sequence 12879, A
162	16.2	67.5	229	3	US-08-750-711-6	Sequence 6, Appl	235	15.8	65.8	305491	4	US-09-949-016-17505	Sequence 17505, A
163	16.2	67.5	549	4	US-09-252-991A-10710	Sequence 10710, A	236	15.6	65.0	140	4	US-09-513-999C-8757	Sequence 8757, App
164	16.2	67.5	567	4	US-09-902-540-7734	Sequence 7734, App	237	15.6	65.0	279	3	US-09-328-869-13	Sequence 13, Appl
165	16.2	67.5	636	4	US-09-252-991A-10456	Sequence 10456, A	238	15.6	65.0	279	3	US-09-628-869-13	Sequence 13, Appl
166	16.2	67.5	1728	4	US-09-902-540-6495	Sequence 6495, App	239	15.6	65.0	279	3	US-09-902-540-3040	Sequence 3040, App
167	16.2	67.5	1953	4	US-09-902-540-6495	Sequence 6495, App	240	15.6	65.0	289	3	US-09-328-869-12	Sequence 12, Appl
168	16.2	67.5	2355	4	US-09-252-991A-3804	Sequence 3804, App	241	15.6	65.0	289	3	US-09-628-869-12	Sequence 12, Appl
169	16.2	67.5	2732	4	US-09-252-991A-3845	Sequence 3845, App	242	15.6	65.0	343	4	US-09-471-296-1606	Sequence 1606, App
170	16.2	67.5	4167	4	US-09-902-540-465	Sequence 465, App	243	15.6	65.0	343	4	US-09-513-999C-8365	Sequence 8365, App
171	16.2	67.5	4935	2	US-08-631-097-3	Sequence 3, Appl	244	15.6	65.0	363	4	US-09-513-999C-12088	Sequence 12088, A
172	16.2	67.5	5886	3	US-08-810-712-9	Sequence 9, Appl	245	15.6	65.0	403	4	US-09-513-999C-31742	Sequence 31742, A
173	16.2	67.5	9818	4	US-09-902-540-987	Sequence 987, App	246	15.6	65.0	437	4	US-09-513-999C-10616	Sequence 10616, A

C 247	15.6	65.0	441	4	US-09-270-767-1759	Sequence 3759, Ap
C 248	15.6	65.0	441	4	US-09-270-767-19041	Sequence 19041, A
C 249	15.6	65.0	450	4	US-09-252-991A-6540	Sequence 6540, Ap
C 250	15.6	65.0	450	4	US-09-252-991A-8407	Sequence 8407, Ap
C 251	15.6	65.0	452	4	US-09-513-999C-1145	Sequence 1145, Ap
C 252	15.6	65.0	458	3	US-09-060-756-551	Sequence 551, App
C 253	15.6	65.0	458	4	US-09-670-314-551	Sequence 551, App
C 254	15.6	65.0	462	4	US-09-252-991A-13388	Sequence 13388, A
C 255	15.6	65.0	467	4	US-09-621-976-1973	Sequence 1973, Ap
C 256	15.6	65.0	467	4	US-09-621-976-17391	Sequence 17391, A
C 257	15.6	65.0	471	4	US-09-513-999C-36158	Sequence 36158, A
C 258	15.6	65.0	516	4	US-09-252-991A-508	Sequence 508, App
C 259	15.6	65.0	537	4	US-09-252-991A-5627	Sequence 5627, App
C 260	15.6	65.0	547	4	US-09-513-999C-3905	Sequence 3905, Ap
C 261	15.6	65.0	573	4	US-09-902-540-4384	Sequence 4384, Ap
C 262	15.6	65.0	600	4	US-09-894-844-41	Sequence 41, Appl
C 263	15.6	65.0	601	4	US-09-949-016-29248	Sequence 29248, A
C 264	15.6	65.0	601	4	US-09-949-016-45231	Sequence 45231, A
C 265	15.6	65.0	601	4	US-09-949-016-58791	Sequence 58791, A
C 266	15.6	65.0	601	4	US-09-949-016-58792	Sequence 58792, A
C 267	15.6	65.0	601	4	US-09-949-016-62891	Sequence 62891, A
C 268	15.6	65.0	601	4	US-09-949-016-202923	Sequence 202923, A
C 269	15.6	65.0	624	4	US-09-252-991A-6551	Sequence 6551, App
C 270	15.6	65.0	648	4	US-09-107-433-1573	Sequence 1573, App
C 271	15.6	65.0	672	4	US-09-252-991A-9502	Sequence 9502, App
C 272	15.6	65.0	672	4	US-09-949-016-465	Sequence 465, App
C 273	15.6	65.0	677	4	US-09-949-016-4472	Sequence 4472, App
C 274	15.6	65.0	678	4	US-09-902-540-9065	Sequence 9065, App
C 275	15.6	65.0	723	4	US-09-252-991A-13839	Sequence 13839, A
C 276	15.6	65.0	723	4	US-09-902-540-7404	Sequence 7404, App
C 277	15.6	65.0	744	4	US-09-489-039A-6647	Sequence 6647, App
C 278	15.6	65.0	765	4	US-09-252-991A-6511	Sequence 6511, App
C 279	15.6	65.0	788	3	US-09-142-623-10	Sequence 10, Appl
C 280	15.6	65.0	795	4	US-09-902-540-6542	Sequence 6542, App
C 281	15.6	65.0	822	4	US-09-252-991A-8940	Sequence 8940, App
C 282	15.6	65.0	837	4	US-09-902-540-3467	Sequence 3467, App
C 283	15.6	65.0	843	4	US-09-902-540-8507	Sequence 8507, App
C 284	15.6	65.0	845	4	US-09-902-540-8861	Sequence 8861, App
C 285	15.6	65.0	864	4	US-09-252-991A-6520	Sequence 6520, App
C 286	15.6	65.0	867	4	US-09-252-991A-13023	Sequence 13023, A
C 287	15.6	65.0	875	4	US-09-902-540-3349	Sequence 3349, App
C 288	15.6	65.0	894	4	US-09-252-991A-11968	Sequence 11968, A
C 289	15.6	65.0	897	4	US-09-949-016-2619	Sequence 2619, App
C 290	15.6	65.0	897	4	US-09-949-016-2900	Sequence 2900, App
C 291	15.6	65.0	908	3	US-08-660-645A-4	Sequence 4, Appl
C 292	15.6	65.0	908	3	US-09-298-718-4	Sequence 4, Appl
C 293	15.6	65.0	908	3	US-09-546-969-4	Sequence 4, Appl
C 294	15.6	65.0	908	4	US-09-547-267-4	Sequence 4, Appl
C 295	15.6	65.0	918	4	US-09-252-991A-8355	Sequence 8355, App
C 296	15.6	65.0	921	4	US-09-252-991A-12978	Sequence 12978, A
C 297	15.6	65.0	954	4	US-09-489-039A-2597	Sequence 2597, App
C 298	15.6	65.0	975	4	US-09-252-991A-12499	Sequence 12499, A
C 299	15.6	65.0	984	3	US-08-459-046-1	Sequence 1, Appl
C 300	15.6	65.0	984	3	US-08-513-974B-57	Sequence 57, Appl

## ALIGNMENTS

RESULT 1  
 US-09-252-991A-10060  
 ; Sequence 10060, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 10060  
 ; LENGTH: 669  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-10060  
 Query Match 78.3%; Score 18.8; DB 4; Length 669;  
 Best Local Similarity 70.8%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Cy 1 GTCTGTGCGCGTACGCGSAC 24  
 Db 252 GTCTGTGCGCGCGCACCGCAC 275

RESULT 2  
 US-09-252-991A-9773/c  
 ; Sequence 9773, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 9773  
 ; LENGTH: 810  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-9773

Query Match 78.3%; Score 18.8; DB 4; Length 810;  
 Best Local Similarity 70.8%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Cy 1 GTCTGTGCGCGTACGCGSAC 24  
 Db 634 GTCTGTGCGCGCACCGCAC 611

RESULT 3  
 US-09-252-991A-8162/c  
 ; Sequence 8162, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 8162  
 ; LENGTH: 1338  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-8162

Query Match 78.3%; Score 18.8; DB 4; Length 1338;  
 Best Local Similarity 70.8%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

QY      1 GTCCTGTTGGCCSGTSAACSGSAC 24
          ||||:||||:||||:||||
Db      999 GTCCGGCTGGCCGGTCAACCCAC 976

```

```

US-09-902-540-4394/C
: Sequence 4394, Application US/09902540
: Patent No. 6833447
: GENERAL INFORMATION:
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Wiegand, Roger C.
: TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
: FILE REFERENCE: 38-10(115849)B
: CURRENT APPLICATION NUMBER: US/09/902,540
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: 60/217, 883
: PRIOR FILING DATE: 2000-07-10
: NUMBER OF SEQ ID NOS: 16825
: SEQ ID NO 4394
: LENGTH: 2403
: TYPE: DNA
: ORGANISM: Myxococcus xanthus
US-09-902-540-4394

```

Query Match	78.3%;	Score 18.8;	DB 4;	Length 2403;
Best Local Similarity	70.8%;	Pred. No. 1.2e+02;		
Matches 17;	Conservative 5;	Mismatches 2;	Indels 0;	Gaps 0

QY 1 GTCTGTGGCCSGTSACSCGSAC 24  
|||:|:|:|:|:|:|  
Db 777 GTCTGCACGCCCGTCACCCGCAC 754

```

RESULT 5
US-09-252-991A-9867/c
; Sequence 9867, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9867
; LENGTH: 4248
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9867

```

Query Match	78.3%	Score 18.8;	DB 4;	Length 4248;
Best Local Similarity	70.8%;	Pred. No.1.2e+02;		
Matches 17; Conservative	5;	Mismatches 2;	Indels 0;	Gaps 0

QY 1 GTCTGCTGGCCSGTACSCGSAC 24  
|||:|:|:|:|:|:|  
Db 975 GTCTGCTTGGCCGCCACCCGCAC 952

RESULT 6  
US-08-804-227C-13  
; Sequence 13, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

Query Match	78.3%;	Score 18.8;	DB 2;	Length 13987;
Best Local Similarity	70.8%;	Pred. No. 1.2e+02;		
Matches 17; Conservative	5;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	1	GTCTGTGTGGCCSGTSACSGSAC	24
		:           :	
Db	8324	GACCTGTGTGGCCGTTGACGGGCAC	8347

```

RESULT 7
US-09-902-540-1217
; Sequence 1217, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10 (15849) B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1217
; LENGTH: 29272
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1217

```

Query Match	78.3%;	Score 18.8;	DB 4;	Length 29272;
Best Local Similarity	70.8%;	Pred. No. 1.2e+02;		



Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCTGTTGGCCGCTGACGCGCAC 24  
Db 20173 GTCCTGCGACGCCGCTGACGCGCAC 20196

RESULT 8  
US-08-804-227C-7  
Sequence 7, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhse, Stuart A.  
APPLICANT: Rostek, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-227C-7

Query Match 78.3%; Score 18.8; DB 2; Length 4437;  
Best local Similarity 70.8%; Pred. No. 1.2e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCTGTTGGCCGCTGACGCGCAC 24  
Db 8339 GACCTGTGGCCGCTGACGCGCAC 8362

RESULT 9  
US-08-804-198-1  
Sequence 1, Application US/08804198  
Patent No. 5945320  
GENERAL INFORMATION:

APPLICANT: Burgett, Stanley G.  
APPLICANT: Kuhse, Stuart A.  
APPLICANT: Rao, Nagendra R.  
APPLICANT: Richardson, Mark A.  
APPLICANT: Rostek, Paul R., Jr.  
TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PAUL R. CANTRELL 1138  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,198  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-198-1

Query Match 78.3%; Score 18.8; DB 2; Length 4437;  
Best local Similarity 70.8%; Pred. No. 1.2e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCTGTTGGCCGCTGACGCGCAC 24  
Db 8339 GACCTGTGGCCGCTGACGCGCAC 8362

RESULT 10  
US-09-902-540-3008  
Sequence 3008, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.

```

; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,893
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3008
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; US-09-902-540-3008

Query Match          75.0%; Score 18; DB 4; Length 432;
Best Local Similarity 75.0%; Pred. No. 2.6e+02;
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCTGTGGCCGCGTSAACGCGS 22
    |||:|||||:|||||:
DB 75 CCTGTGGCCGCGTSAACGCGC 94

RESULT 11
US-09-902-540-636/c
; Sequence 636, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 636
; LENGTH: 3092
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; US-09-902-540-636

Query Match          75.0%; Score 18; DB 4; Length 3092;
Best Local Similarity 75.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCTGTGGCCGCGTSAACGCGS 22
    |||:|||||:|||||:
DB 3017 CCTGTGGCCGCGTSAACGCGC 2998

RESULT 12
US-09-252-991A-1308/c
; Sequence 1308, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1308
```

```

; LENGTH: 375
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-1308

Query Match          74.2%; Score 17.8; DB 4; Length 375;
Best Local Similarity 69.6%; Pred. No. 3.1e+02;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGGCCGCGTSAACGCGSA 23
    |||:|||||:|||||:
DB 117 GTCTGTGGCCGCGTSAACGCGGA 95

RESULT 13
US-09-252-991A-1157
; Sequence 1157, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1157
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-1157

Query Match          74.2%; Score 17.8; DB 4; Length 681;
Best Local Similarity 69.6%; Pred. No. 3.1e+02;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGGCCGCGTSAACGCGSA 23
    |||:|||||:|||||:
DB 610 GTCTGTGGCCGCGTSAACGCGGA 632

RESULT 14
US-09-252-991A-5982
; Sequence 5982, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5982
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-5982

Query Match          72.5%; Score 17.4; DB 4; Length 492;
Best Local Similarity 71.4%; Pred. No. 4.5e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCTGTGGCCGCGTSAACGCGS 22
    |||:|||||:|||||:
```

Db 232 TCCTGCTGCGCATGACCCGC 252

RESULT 15

US-09-252-991A-339/c  
Sequence 339, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

SEQ ID NO 339

LENGTH: 654

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-339

Query Match 72.5%; Score 17.4; DB 4; Length 654;  
Best Local Similarity 71.4%; Pred. No. 4.5e+02;

Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 548 TCCTGCTGCGCGGTGCGCGG 528

2 TCCTGCTGCGCGGTGCGCGG 22

RESULT 16

US-09-252-991A-330  
Sequence 330, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

SEQ ID NO 330

LENGTH: 759

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-330

Query Match 72.5%; Score 17.4; DB 4; Length 759;  
Best Local Similarity 71.4%; Pred. No. 4.5e+02;

Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 617 TCCTGCTGCGCGGTGCGCGG 637

2 TCCTGCTGCGCGGTGCGCGG 22

RESULT 17

US-09-489-039A-3425  
Sequence 3425, Application US/09489039A  
Patent No. 6610836

GENERAL INFORMATION:

APPLICANT: Gary Breton et. al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001

CURRENT APPLICATION NUMBER: US/09/489,039A

CURRENT FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: US 60/117,747

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 3425

LENGTH: 843

TYPE: DNA

ORGANISM: Klebsiella pneumoniae

US-09-489-039A-3425

Query Match 72.5%; Score 17.4; DB 4; Length 843;  
Best Local Similarity 71.4%; Pred. No. 4.5e+02;

Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 337 CTGCTGCGCGATGACCCGAC 357

4 CTGCTGCGCGATGACCCGAC 24

RESULT 18

US-09-252-991A-319  
Sequence 319, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

SEQ ID NO 319

LENGTH: 1026

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

FEATURE:

NAME/KEY: unsure

LOCATION: (962)

OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.

US-09-252-991A-319

Query Match 72.5%; Score 17.4; DB 4; Length 1026;  
Best Local Similarity 71.4%; Pred. No. 4.5e+02;

Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 78 TCCTGCTGCGCGGTGCGCGG 98

2 TCCTGCTGCGCGGTGCGCGG 22

RESULT 19

5212296-17  
Patent No. 5212296  
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LIETO, KENNETH  
J.; O'KEEFE, DANIEL P.; JOWER, CHARLES A.; ROMESSER, JAMES A.  
TEPPERMAN, JAMES M.

TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING  
CYTOCHROMES

NUMBER OF SEQUENCES: 19

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/569,781

FILING DATE: 23-AUG-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 464,499

FILING DATE: 12-JAN-1990

APPLICATION NUMBER: 405,605

FILING DATE: 11-SEP-1989

```

;SEQ ID NO:17
; LENGTH: 1212
5212296-17

Query Match      72.5%   Score 17.4; DB 6; Length 1212;
Best Local Similarity 71.4%; Pred. No. 4.5e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCCTGATGGCCSGTSACSCGS 22
       |||||:::||:::||::||:
Db      148 TCGTGTGCTGCTGTGACGCCG 168

RESULT 20
5212296-17
Patent No. 5212296
APPLICANT: DEAN, CAROLINE,HARDER, PATRICIA A.,LETO, KENNETH
J., O'KEEFE, DANIEL P.,OMER, CHARLES A.,ROMESSER, JAMES A.
TEPPERMAN, JAMES M.
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
CYTOCHROMES
NUMBER OF SEQUENCES: 19
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/569,781
FILING DATE: 23-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 464,499
FILING DATE: 12-JAN-1990
APPLICATION NUMBER: 405,605
FILING DATE: 11-SEP-1989
;SEQ ID NO:17:
; LENGTH: 1212
5212296-17

Query Match      72.5%   Score 17.4; DB 6; Length 1212;
Best Local Similarity 71.4%; Pred. No. 4.5e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCCTGATGGCCSGTSACSCGS 22
       |||||:::||:::||::||:
Db      148 TCGTGTGCTGCTGTGACGCCG 168

RESULT 21
US-09-252-991A-5965
Sequence 5965, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: ABRUINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 5965
LENGTH: 1578
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5965

Query Match      72.5%   Score 17.4; DB 4; Length 1578;
Best Local Similarity 71.4%; Pred. No. 4.4e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCCTGATGGCCSGTSACSCGS 22
       |||||:::||:::||::||:
Db      434 TCGTGTGCTGCTGTGACGCCG 454
```

```

RESULT 22
US-09-252-991A-11053
; Sequence 11053, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11053
; LENGTH: 1827
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11053

Query Match          72.5%; Score 17.4; DB 4; Length 1827;
Best Local Similarity 71.4%; Pred. No. 4,4e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY      3 CCTGCTGCGCGGTGACGCGSA 23
      |||:|||||:|:||:|:|
Db      1660 CCTGCTGCGCGGCGACCGCA 1640

RESULT 23
US-09-252-991A-11201/C
; Sequence 11201, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11201
; LENGTH: 1929
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11201

Query Match          72.5%; Score 17.4; DB 4; Length 1929;
Best Local Similarity 71.4%; Pred. No. 4,4e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY      3 CCTGCTGCGCGGTGACGCGSA 23
      |||:|||||:|:||:|:|
Db      1660 CCTGCTGCGCGGCGACCGCA 1640

RESULT 24
US-09-252-991A-10982
; Sequence 10982, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A

```

CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 10982  
LENGTH: 1962  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10982

Query Match 72.5%; Score 17.4; DB 4; Length 1962;  
Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCTGTGGCCGCGTACCGCGA 23  
Db 289 CCTGTGGCCGCGCGCGCGA 309

RESULT 25  
5212296-8  
PATENT NO. 5212296  
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH  
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.  
TEPPERMAN, JAMES M.  
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING  
CYTOCHROMES  
NUMBER OF SEQUENCES: 19  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/569,781  
FILING DATE: 23-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 464,499  
FILING DATE: 12-JAN-1990  
APPLICATION NUMBER: 405,605  
FILING DATE: 11-SEP-1989  
SEQ ID NO: 8  
LENGTH: 1998

Query Match 72.5%; Score 17.4; DB 6; Length 1998;  
Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCCGTGTGGCCGCGTACCGCG 22  
Db 342 TCCGTGTGGCGTGTGACCGCG 362

RESULT 26  
5212296-8  
PATENT NO. 5212296  
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH  
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.  
TEPPERMAN, JAMES M.  
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING  
CYTOCHROMES  
NUMBER OF SEQUENCES: 19  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/569,781  
FILING DATE: 23-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 464,499  
FILING DATE: 12-JAN-1990  
APPLICATION NUMBER: 405,605  
FILING DATE: 11-SEP-1989  
SEQ ID NO: 8  
LENGTH: 1998

Query Match 72.5%; Score 17.4; DB 6; Length 1998;

Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCCGTGTGGCCGCGTACCGCG 22  
Db 342 TCCGTGTGGCGTGTGACCGCG 362

RESULT 27  
US-09-252-991A-5908/c  
Sequence 5908, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: MARC J. RUBENFIELD et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 5908  
LENGTH: 2469  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5908

Query Match 72.5%; Score 17.4; DB 4; Length 2469;  
Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCCGTGTGGCCGCGTACCGCG 22  
Db 786 TCCGTGTGGCGGATACCGCG 766

RESULT 28  
US-09-252-991A-5922/c  
Sequence 5922, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: MARC J. RUBENFIELD et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 5922  
LENGTH: 3135  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5922

Query Match 72.5%; Score 17.4; DB 4; Length 3135;  
Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCCGTGTGGCCGCGTACCGCG 22  
Db 2132 TCCGTGTGGCGGATACCGCG 2112

RESULT 29  
US-09-902-540-8359  
Sequence 8359, Application US/09902540

```
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8359
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-8359
```

```
Query Match          71.7%; Score 17.2; DB 4; Length 444;
Best Local Similarity 66.7%; Pred. No. 5.4e+02;
Matches 16; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1  GTCTGTGTGCGCGTGTACGCGCTC 24
Db      270 GCCCTGTGTGCGCGTGTACGCGCTC 293
```

```
RESULT 30
US-09-252-991A-6007/C
; Sequence 6007, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6007
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6007
```

```
Query Match          71.7%; Score 17.2; DB 4; Length 585;
Best Local Similarity 66.7%; Pred. No. 5.4e+02;
Matches 16; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1  GTCTGTGTGCGCGTGTACGCGSAC 24
Db      552 GCCGTCTGTGCGCGTGTACCGGAC 529
```

Search completed: July 20, 2005, 16:48:37  
Job time : 131.714 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 15:46:36 ; Search time 2983.43 Seconds  
(without alignments)  
306.206 Million cell updates/sec

Title: US-10-017-471B-2

Perfect score: 24  
Sequence: 1 gtcctgctgcccgcacacgacgac 24

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 300 summaries

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hlc: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_g981: \*  
9: gb\_g982: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.4	80.8	1059	2	BB869584 601445696
C 2	18.8	78.3	132	7	CNT719278 E0781F12-
C 3	18.8	78.3	247	2	BB869974 BB869974
C 4	18.8	78.3	315	7	CK929651 p5tmgc_00
C 5	18.8	78.3	338	2	BB868549 BB868549
C 6	18.8	78.3	396	5	BY280059 BY280059
C 7	18.8	78.3	428	7	CK929166 p5tmgc_00
C 8	18.8	78.3	432	2	BB863614 BB863614
C 9	18.8	78.3	447	2	BB839789 BB839789
C 10	18.8	78.3	481	1	AA036131 m170h10.r
C 11	18.8	78.3	510	6	CA879141 K0972B03-
C 12	18.8	78.3	516	6	CB060700 AMGNNUC:N
C 13	18.8	78.3	526	6	CB060700 AMGNNUC:N
C 14	18.8	78.3	538	7	CK928924 p5tmgc_00
C 15	18.8	78.3	551	2	BR462782 UT-M-CGDP
C 16	18.8	78.3	577	8	A2417416 IM0193N06
C 17	18.8	78.3	610	2	BB642652 BB642652
C 18	18.8	78.3	619	6	BY709847 BY709847
C 19	18.8	78.3	641	5	BTU01308 UT-M-EXO-
C 20	18.8	78.3	672	7	CNT721004 E0814B04-
C 21	18.8	78.3	717	4	BT732637 603355581
C 22	18.8	78.3	760	4	BZ550710 pacel-60_
C 23	18.8	78.3	780	1	AU080793 AU080793
C 24	18.8	78.3	782	8	BZ565923 pacel-164

C 25	18.8	78.3	793	1	AU080514
C 26	18.8	78.3	907	8	CNS077KA
C 27	18.8	78.3	909	8	BZ568254
C 28	18.8	78.3	927	9	CNS0795W
C 29	18.8	78.3	1024	9	CNS0795W
C 30	18.8	78.3	1325	3	AK010405
C 31	18.4	76.7	148	2	BF748536
C 32	18.4	76.7	697	2	BE870471
C 33	18.4	76.7	979	2	BF309584
C 34	18.2	75.8	300	6	BY694846
C 35	18.2	75.8	598	4	BM338416
C 36	18.2	75.8	666	7	CK941631
C 37	18.2	75.8	700	4	BM348799
C 38	18.2	75.8	900	5	BUS13470
C 39	18.2	75.0	709	8	BZ755081
C 40	18.2	75.0	744	9	CG268813
C 41	18.2	75.0	802	8	BZ755082
C 42	18.2	75.0	813	9	CG268801
C 43	18.2	75.0	826	9	CG368155
C 44	18.2	75.0	888	9	CG720046
C 45	18.2	75.0	893	9	CG346247
C 46	18.2	75.0	914	9	CG171982
C 47	18.2	75.0	960	9	CG368165
C 48	17.8	74.2	357	5	BY008993
C 49	17.8	74.2	424	5	BM473192
C 50	17.8	74.2	455	5	BM449237
C 51	17.8	74.2	485	7	CK929214
C 52	17.8	74.2	497	9	CG344719
C 53	17.8	74.2	524	4	BM337458
C 54	17.8	74.2	529	1	AV885559
C 55	17.8	74.2	555	9	CG217868
C 56	17.8	74.2	585	9	CE294296
C 57	17.8	74.2	590	1	AV839300
C 58	17.8	74.2	626	2	AW739943
C 59	17.8	74.2	629	5	BU040352
C 60	17.8	74.2	629	5	BM070493
C 61	17.8	74.2	630	9	CL981750
C 62	17.8	74.2	671	4	BG767585
C 63	17.8	74.2	683	5	BM172508
C 64	17.8	74.2	699	9	CE806491
C 65	17.8	74.2	705	5	BM052673
C 66	17.8	74.2	707	5	BM374464
C 67	17.8	74.2	723	5	BM631806
C 68	17.8	74.2	723	5	BM105843
C 69	17.8	74.2	723	7	CF866166
C 70	17.8	74.2	728	9	CL552097
C 71	17.8	74.2	746	9	AG549784
C 72	17.8	74.2	747	5	BM377235
C 73	17.8	74.2	749	5	BM404530
C 74	17.8	74.2	754	5	BM023548
C 75	17.8	74.2	758	7	CK132546
C 76	17.8	74.2	773	5	BP019079
C 77	17.8	74.2	774	5	BM083721
C 78	17.8	74.2	778	6	CB896124
C 79	17.8	74.2	782	5	BM379174
C 80	17.8	74.2	820	5	BO708394
C 81	17.8	74.2	947	8	BZ567311
C 82	17.8	74.2	1017	9	CNS0273VK
C 83	17.8	74.2	1066	9	CL082900
C 84	17.8	74.2	1228	8	BZ572830
C 85	17.8	74.2	1330	8	BZ572833
C 86	17.6	73.3	321	8	CK919171
C 87	17.6	73.3	421	7	CK929813
C 88	17.6	73.3	557	2	AM012945
C 89	17.6	73.3	583	4	BM575120
C 90	17.6	73.3	633	4	CNS031ZB
C 91	17.6	73.3	943	1	AA015262
C 92	17.6	73.3	985	4	BI252388
C 93	17.6	73.3	1026	5	BO653212
C 94	17.6	73.3	1078	9	BM549RDB
C 95	17.4	72.5	239	2	BF817181
C 96	17.4	72.5	363	9	CE783170
C 97	17.4	72.5	542	8	BM782222

AA038200 T7 end of	AA080514
BZ568254 pacel-164	AA080514
AL435592 T3 end of	AA080514
AL434874 T3 end of	AA080514
AK010405 Mus muscu	AA080514
BF748536 RC2-BN040	AA080514
BE870471 601447033	AA080514
BF309584 601891902	AA080514
BY694846 BY694846	AA080514
BM338416 MEST226-F	AA080514
CK941631 4065193 B	AA080514
BM348799 MEST301-C	AA080514
BUS13470 AGENCOURT	AA080514
BZ755081 PUPBU91TB	AA080514
CG268813 CGMB035TV	AA080514
BZ755082 PUPBU91TB	AA080514
CG268801 OGBRC35TH	AA080514
CG368155 OGBRC35TH	AA080514
CG720046 OGBRC35TH	AA080514
CG346247 OGBRC35TH	AA080514
CG171982 PUGO82TD	AA080514
CG368165 OGBRC35TV	AA080514
BY008993 BY008993	AA080514
BM473192 BM473192	AA080514
BM449237 BM449237	AA080514
CK929214 p5tmgc_00	AA080514
CG344719 OGBRC35TH	AA080514
BM337458 MEST206-H	AA080514
AV885559 AV885559	AA080514
CG217868 OGBRC35TH	AA080514
CE294296 ligr-g98-	AA080514
AV839300 AV839300	AA080514
AW739943 BR100456	AA080514
BU040352 603483802	AA080514
BM070493 BM070493	AA080514
CL981750 OGBRC35TH	AA080514
BG767585 602741743	AA080514
BM172508 BM172508	AA080514
CE806491 ligr-g98-	AA080514
BM052673 BM052673	AA080514
BM374464 BM374464	AA080514
BM631806 170006688	AA080514
BM105843 BM105843	AA080514
CF866166 ligr-g98-	AA080514
CL552097 OB_BA009	AA080514
AG549784 Mus muscu	AA080514
BM377235 BM377235	AA080514
BM404530 BM404530	AA080514
BM023548 BM023548	AA080514
CK132546 SB02035A1	AA080514
BP019079 BP019079	AA080514
BM083721 BM083721	AA080514
CB896124 ligr-g98-	AA080514
BM379174 BM379174	AA080514
BO708394 AGENCOURT	AA080514
BZ567311 pacel-164	AA080514
AL185177 Tetracodon	AA080514
CL082900 ISB1-IC2_	AA080514
BZ572830 msh2_2820	AA080514
BZ572833 msh2_2817	AA080514
CK919171 HS-1049-A1-	AA080514
CK929813 p5tmgc_00	AA080514
AM012945 Odc-0045	AA080514
BM575120 Tetracodon	AA080514
AL264800 ligr-g98-	AA080514
AA015262 msh20d10.r	AA080514
BM575120 602953228	AA080514
BO653212 Tetracodon	AA080514
AL303752 Tetracodon	AA080514
BF817181 RC4-CT002	AA080514
CE783170 ligr-g98-	AA080514
BM782222 fzmb011f0	AA080514

C 98	17.4	72.5	553	2	AM498322	AM498322	660046B11	C 171	17.2	71.7	419	7	CO525050	CO525050	3530_1_16
C 99	17.4	72.5	560	4	CE731010	CE731010	l1gr-g8s-	C 172	17.2	71.7	423	7	CK930464	CK930464	p5imgc_00
C 100	17.4	72.5	569	4	BI340535	BI340535	365832 MA	C 173	17.2	71.7	426	7	CK930385	CK930385	p5imgc_00
C 101	17.4	72.5	578	4	BI340506	BI340506	365796 MA	C 174	17.2	71.7	432	7	CK930422	CK930422	p5imgc_00
C 102	17.4	72.5	620	8	BZ708597	BZ708597	OGBAV35TM	C 175	17.2	71.7	434	2	AM374956	MR0-CT006	
C 103	17.4	72.5	643	7	CO521649	CO521649	3530_1_14	C 176	17.2	71.7	436	2	BB671056	BB671056	p5imgc_00
C 104	17.4	72.5	657	7	CO523848	CO523848	3530_1_14	C 177	17.2	71.7	449	7	CK930114	CK930114	p5imgc_00
C 105	17.4	72.5	680	8	BZ564040	BZ564040	pac62-164	C 178	17.2	71.7	450	7	CK930187	CK930187	p5imgc_00
C 106	17.4	72.5	699	8	AY104786	AY104786	zea mayr	C 179	17.2	71.7	450	7	CK930428	CK930428	p5imgc_00
C 107	17.4	72.5	721	8	BZ708588	BZ708588	OGBAV35TC	C 180	17.2	71.7	451	7	CK930640	CK930640	p5imgc_00
C 108	17.4	72.5	791	4	CC372496	CC372496	PUBSR04TD	C 181	17.2	71.7	454	7	CK929187	CK929187	p5imgc_00
C 109	17.4	72.5	821	4	BC320099	BC320099	Zm03_01f0	C 182	17.2	71.7	459	4	B1795700	B1795700	H029C05 E
C 110	17.4	72.5	835	6	CA085587	CA085587	SCUTAM209	C 183	17.2	71.7	460	7	CK929968	CK929968	p5imgc_00
C 111	17.4	72.5	874	2	BF215026	BF215026	60184735B	C 184	17.2	71.7	461	7	CNI181198	CNI181198	SGP169T12
C 112	17.4	72.5	1028	9	CG115513	CG115513	PUID065TB	C 185	17.2	71.7	463	7	CK929790	CK929790	p5imgc_00
C 113	17.4	72.5	1064	9	CNS02073	CNS02073	Al216666	C 186	17.2	71.7	464	7	CK930564	CK930564	p5imgc_00
C 114	17.4	72.5	1064	9	BC030024	BC030024	Hom sapi	C 187	17.2	71.7	466	7	CK929233	CK929233	p5imgc_00
C 115	17.2	71.7	199	2	BB602482	BB602482	Hom sapi	C 188	17.2	71.7	469	7	CK930176	CK930176	p5imgc_00
C 116	17.2	71.7	206	8	BH256468	BH256468	KG03203 D	C 189	17.2	71.7	472	5	BX526676	BX526676	
C 117	17.2	71.7	208	8	BH538521	BH538521	KG03203 D	C 190	17.2	71.7	474	1	AL830986	AL830986	
C 118	17.2	71.7	218	7	CK930931	CK930931	p5imgc_00	C 191	17.2	71.7	475	7	CK929207	CK929207	p5imgc_00
C 119	17.2	71.7	220	7	CK930073	CK930073	p5imgc_00	C 192	17.2	71.7	476	7	CK930749	CK930749	p5imgc_00
C 120	17.2	71.7	239	7	CK929744	CK929744	p5imgc_00	C 193	17.2	71.7	476	5	CO524912	CO524912	3530_1_16
C 121	17.2	71.7	239	7	CK931038	CK931038	p5imgc_00	C 194	17.2	71.7	479	5	BP459026	BP459026	
C 122	17.2	71.7	244	7	CK930396	CK930396	p5imgc_00	C 195	17.2	71.7	479	7	CK928916	CK928916	p5imgc_00
C 123	17.2	71.7	251	7	CK930775	CK930775	p5imgc_00	C 196	17.2	71.7	480	7	CK929740	CK929740	p5imgc_00
C 124	17.2	71.7	251	7	CK930934	CK930934	p5imgc_00	C 197	17.2	71.7	480	7	CK930518	CK930518	p5imgc_00
C 125	17.2	71.7	260	7	CK929060	CK929060	p5imgc_00	C 198	17.2	71.7	481	9	AG257747	Lotus cor	
C 126	17.2	71.7	260	7	CK930516	CK930516	p5imgc_00	C 199	17.2	71.7	484	7	CK929216	CK929216	p5imgc_00
C 127	17.2	71.7	276	7	H27643	H27643	Y1E1h01.T1	C 200	17.2	71.7	485	6	BY597533	BY597533	
C 128	17.2	71.7	277	7	CK930449	CK930449	p5imgc_00	C 201	17.2	71.7	486	7	CK930043	CK930043	p5imgc_00
C 129	17.2	71.7	282	7	CK929198	CK929198	p5imgc_00	C 202	17.2	71.7	487	7	CK929590	CK929590	p5imgc_00
C 130	17.2	71.7	286	7	CK930413	CK930413	p5imgc_00	C 203	17.2	71.7	490	7	CK929527	CK929527	p5imgc_00
C 131	17.2	71.7	292	2	BB208092	BB208092	BB208092	C 204	17.2	71.7	491	7	CK131411	LP02583.3	
C 132	17.2	71.7	297	2	BB360214	BB360214	BB360214	C 205	17.2	71.7	491	7	CK929837	CK929837	p5imgc_00
C 133	17.2	71.7	298	7	CK929949	CK929949	p5imgc_00	C 206	17.2	71.7	491	7	CK930352	CK930352	p5imgc_00
C 134	17.2	71.7	299	7	CK930369	CK930369	p5imgc_00	C 207	17.2	71.7	494	7	CK929494	CK929494	p5imgc_00
C 135	17.2	71.7	300	7	CK929576	CK929576	p5imgc_00	C 208	17.2	71.7	497	7	CK929601	CK929601	p5imgc_00
C 136	17.2	71.7	304	7	CK929850	CK929850	p5imgc_00	C 209	17.2	71.7	497	7	CK930832	CK930832	p5imgc_00
C 137	17.2	71.7	305	7	CO524947	CO524947	3530_1_16	C 210	17.2	71.7	498	6	CA130140	SCVPLR201	
C 138	17.2	71.7	306	7	CP857349	CP857349	p5imgc_00	C 211	17.2	71.7	500	7	CK929718	CK929718	p5imgc_00
C 139	17.2	71.7	308	7	CE853616	CE853616	p5imgc_00	C 212	17.2	71.7	500	7	CK929999	CK929999	p5imgc_00
C 140	17.2	71.7	309	1	AU603499	AU603499	p5imgc_00	C 213	17.2	71.7	505	7	CK929574	CK929574	p5imgc_00
C 141	17.2	71.7	320	7	CK930307	CK930307	p5imgc_00	C 214	17.2	71.7	507	7	CK929440	CK929440	p5imgc_00
C 142	17.2	71.7	322	9	CE553081	CE553081	l1gr-g8s-	C 215	17.2	71.7	507	7	CK929523	CK929523	p5imgc_00
C 143	17.2	71.7	328	7	CK929725	CK929725	p5imgc_00	C 216	17.2	71.7	507	7	CK930438	CK930438	p5imgc_00
C 144	17.2	71.7	337	2	BP335934	BP335934	CM4-CT068	C 217	17.2	71.7	513	1	AV430005	AV430005	
C 145	17.2	71.7	341	2	AM579227	AM579227	PM2-DT004	C 218	17.2	71.7	517	7	CK929585	CK929585	p5imgc_00
C 146	17.2	71.7	349	7	CK929878	CK929878	p5imgc_00	C 219	17.2	71.7	519	7	CK930252	CK930252	p5imgc_00
C 147	17.2	71.7	355	6	BY594589	BY594589	p5imgc_00	C 220	17.2	71.7	521	7	CK929891	CK929891	p5imgc_00
C 148	17.2	71.7	356	7	CK929008	CK929008	p5imgc_00	C 221	17.2	71.7	524	2	AM564100	AM564100	L61_282 E
C 149	17.2	71.7	359	8	BZ293509	BZ293509	CG0368_r1	C 222	17.2	71.7	524	7	CK929640	CK929640	p5imgc_00
C 150	17.2	71.7	365	5	BP087529	BP087529	p5imgc_00	C 223	17.2	71.7	526	7	CK930103	CK930103	p5imgc_00
C 151	17.2	71.7	372	7	CK929938	CK929938	p5imgc_00	C 224	17.2	71.7	527	7	CK928894	CK928894	p5imgc_00
C 152	17.2	71.7	381	7	CK929756	CK929756	p5imgc_00	C 225	17.2	71.7	529	7	CK930722	CK930722	p5imgc_00
C 153	17.2	71.7	383	7	CO191444	CO191444	BC030874.5	C 226	17.2	71.7	530	7	CK930864	CK930864	p5imgc_00
C 154	17.2	71.7	388	2	AM659731	AM659731	97381 MAR	C 227	17.2	71.7	532	2	BE468773	IPHdK-F004	
C 155	17.2	71.7	390	7	CK9299340	CK9299340	p5imgc_00	C 228	17.2	71.7	532	7	CK929810	CK929810	p5imgc_00
C 156	17.2	71.7	390	7	U47733	U47733	HSU47733_Hu	C 229	17.2	71.7	535	1	A1261133	LP05241.3	
C 157	17.2	71.7	396	5	BP086909	BP086909	p5imgc_00	C 230	17.2	71.7	538	7	CK931028	CK931028	p5imgc_00
C 158	17.2	71.7	396	9	CE158344	CE158344	l1gr-g8s-	C 231	17.2	71.7	539	1	A1258473	LP016807.5	
C 159	17.2	71.7	397	7	CK928626	CK928626	p5imgc_00	C 232	17.2	71.7	539	1	AL915020	AL915020	p5imgc_00
C 160	17.2	71.7	399	7	CK929789	CK929789	p5imgc_00	C 233	17.2	71.7	539	7	CK656943	CK656943	LP22811.3
C 161	17.2	71.7	402	7	CP856066	CP856066	p5imgc_00	C 234	17.2	71.7	539	7	CK929836	CK929836	p5imgc_00
C 162	17.2	71.7	402	7	CK930024	CK930024	p5imgc_00	C 235	17.2	71.7	540	7	CK930378	CK930378	p5imgc_00
C 163	17.2	71.7	403	7	CK929486	CK929486	p5imgc_00	C 236	17.2	71.7	541	7	CK928949	CK928949	p5imgc_00
C 164	17.2	71.7	404	5	BY286951	BY286951	p5imgc_00	C 237	17.2	71.7	543	2	BE419770	WMS017.C1	
C 165	17.2	71.7	409	7	CK929501	CK929501	p5imgc_00	C 238	17.2	71.7	544	1	AV626769	AV626769	p5imgc_00
C 166	17.2	71.7	409	7	CK930143	CK930143	p5imgc_00	C 239	17.2	71.7	544	7	CK929681	CK929681	p5imgc_00
C 167	17.2	71.7	413	7	CK930476	CK930476	p5imgc_00	C 240	17.2	71.7	544	7	CK930548	CK930548	p5imgc_00
C 168	17.2	71.7	413	7	CK930827	CK930827	p5imgc_00	C 241	17.2	71.7	548	2	BB013997	BB013997	125554 MA
C 169	17.2	71.7	417	7	CK930196	CK930196	p5imgc_00	C 242	17.2	71.7	548	7	CK930094	CK930094	p5imgc_00
C 170	17.2	71.7	419	7	CK930313	CK930313	p5imgc_00	C 243	17.2	71.7	549	4	BI344979	BI344979	373645 MA



C 244	17.2	71.7	549	7	CK930091	CK930091 p5rmgc_00
C 245	17.2	71.7	555	7	CK929948	CK929948 p5rmgc_00
C 246	17.2	71.7	555	7	CK930699	CK930699 p5rmgc_00
C 247	17.2	71.7	556	7	CK929021	CK929021 p5rmgc_00
C 248	17.2	71.7	557	6	CD488880	CD488880 T15_F07_T
C 249	17.2	71.7	562	7	CK929425	CK929425 p5rmgc_00
C 250	17.2	71.7	563	7	CK930008	CK930008 p5rmgc_00
C 251	17.2	71.7	567	7	CK929283	CK929283 p5rmgc_00
C 252	17.2	71.7	570	7	CK930365	CK930365 p5rmgc_00
C 253	17.2	71.7	571	7	CF181700	CF181700 818615 MA
C 254	17.2	71.7	572	7	CK929566	CK929566 p5rmgc_00
C 255	17.2	71.7	573	1	AU289764	AU289764 AU289764
C 256	17.2	71.7	574	6	CA287868	CA287868 SCSFSD206
C 257	17.2	71.7	575	7	CK929325	CK929325 p5rmgc_00
C 258	17.2	71.7	576	7	CK929866	CK929866 p5rmgc_00
C 259	17.2	71.7	577	7	CK930079	CK930079 p5rmgc_00
C 260	17.2	71.7	579	7	CK929895	CK929895 p5rmgc_00
C 261	17.2	71.7	581	9	CE411108	CE411108 l1gr-gms-
C 262	17.2	71.7	581	6	CD810583	CD810583 1336 Ling
C 263	17.2	71.7	582	4	BM076672	BM076672 T1EST-A08
C 264	17.2	71.7	582	7	CK930208	CK930208 p5rmgc_00
C 265	17.2	71.7	585	7	CK931090	CK931090 p5rmgc_00
C 266	17.2	71.7	587	7	CF850440	CF850440 p5MA014XF
C 267	17.2	71.7	587	7	CK661879	CK661879 Lp21657.5
C 268	17.2	71.7	587	7	CK929840	CK929840 p5rmgc_00
C 269	17.2	71.7	589	5	BQ464452	BQ464452 HF02F05r
C 270	17.2	71.7	589	5	AG271420	AG271420 Cyanidlos
C 271	17.2	71.7	594	7	CK929692	CK929692 p5rmgc_00
C 272	17.2	71.7	595	4	BG368479	BG368479 HVSME1001
C 273	17.2	71.7	595	7	CK929242	CK929242 p5rmgc_00
C 274	17.2	71.7	598	7	CK929860	CK929860 p5rmgc_00
C 275	17.2	71.7	601	7	CK930633	CK930633 p5rmgc_00
C 276	17.2	71.7	602	7	CK929936	CK929936 p5rmgc_00
C 277	17.2	71.7	604	7	CK930067	CK930067 p5rmgc_00
C 278	17.2	71.7	611	7	CK930981	CK930981 p5rmgc_00
C 279	17.2	71.7	625	7	CK930981	CK930981 p5rmgc_00
C 280	17.2	71.7	629	8	BH840371	BH840371 LMC20001
C 281	17.2	71.7	630	4	BT74091	BT74091 1031069A1
C 282	17.2	71.7	630	4	CK930334	CK930334 p5rmgc_00
C 283	17.2	71.7	632	9	CG205632	CG205632 PULC801TD
C 284	17.2	71.7	634	9	CG840062	CG840062 Ynhw0676
C 285	17.2	71.7	636	4	BT548787	BT548787 603189188
C 286	17.2	71.7	640	4	BG518790	BG518790 602578382
C 287	17.2	71.7	642	4	BJ649918	BJ649918 BJ649918
C 288	17.2	71.7	645	7	CK930223	CK930223 p5rmgc_00
C 289	17.2	71.7	645	8	BZ329818	BZ329818 hv8b05_b
C 290	17.2	71.7	654	5	BQ812592	BQ812592 1030031B0
C 291	17.2	71.7	667	7	CK303145	CK303145 SB0201B2
C 292	17.2	71.7	677	7	CK940602	CK940602 4064046 B
C 293	17.2	71.7	680	9	CG242034	CG242034 OGV248TH
C 294	17.2	71.7	684	7	CK930290	CK930290 p5rmgc_00
C 295	17.2	71.7	695	6	CD924088	CD924088 G750_111H
C 296	17.2	71.7	703	5	BU452052	BU452052 603218379
C 297	17.2	71.7	713	5	BU213716	BU213716 603756629
C 298	17.2	71.7	714	7	CV104150	CV104150 AGENCOCURT
C 299	17.2	71.7	716	7	CK940880	CK940880 4064430 B
C 300	17.2	71.7	719	5	BU137350	BU137350 603007751

## ALIGNMENTS

RESULT 1  
BE869584/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BE869584  
BE869584  
BE869584.1  
EST.  
Homo sapiens  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1059 bp  
mRNA  
linear  
EST 20-OCT-2000  
60144569661 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3849863 5',  
mRNA sequence.  
GI:10318464

REFERENCE  
1. (bases 1 to 1059)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LLM9567 row: p column: 24  
High quality sequence stop: 719.  
Location/Qualifiers

FEATURES  
SOURCE  
1. 1059  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3849863"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_65"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.8 kb. Library constructed by Life Technologies."

## ORIGIN

Query Match 80.8%; Score 19.4; DB 2; Length 1059;  
Best Local Similarity 73.9%; Pred. No. 1.5e+03;  
Matches 17; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCTGTGGCGCGTACCGSAC 24  
DB 967 TCTGTGGCGCGTACCGTAC 945

RESULT 2  
CN719278/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CN719278  
E0781F12-5 NIA Mouse four-cell-Embryo cDNA library (long) Mus musculus cDNA clone NIA:E0781F12 IMAGE:30903815 5', mRNA sequence.  
CN719278  
CN719278.1 GI:47488663  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathia; Muridae; Murine; Mus.  
1 (bases 1 to 132)  
Sharov,A.A., Piao,Y., Matoba,R., Dudekula,D.B., Qian,Y., VanBuren,V., Falco,G., Martin,P.R., Stagg,C.A., Basseys,U.C., Wang,Y., Carter,M.G., Hamatani,T., Albe,K., Akutsu,H., Sharova,L., Tanaka,T.S., Kimber,W.L., Yoshikawa,T., Jaradat,S.A., Pantano,S., Nagaraja,R., Boheler,K.R., Taub,D., Hodess,R.J., Longo,D.L., Schlessinger,D., Keller,U., Klotz,E., Kelsoe,G., Umezawa,A., Vesecovi,A.L., Rossant,J., Kunath,T., Hogan,B.L., Curci,A., D'Urso,M., Klesio,U., Hide,W. and Ko,M.S.  
Transcriptome analysis of mouse stem cells and early embryos  
PLoS Biol. 1 (3), 410-419 (2003)  
Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: [cdna@gsun-grc.nia.nih.gov](mailto:cdna@gsun-grc.nia.nih.gov)  
Plate: E0781 row: F column: 12  
Seq primer: M13 Reverse  
High quality sequence stop: 132  
POLYA=No.  
Location/Qualifiers

## FEATURES

## source

1. 132  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /db\_xref="taxon:10090"  
 /clone="NIA:E0781P12 IMAGE:30903815"  
 /tissue\_type="4-cell stage embryo"  
 /dev\_stage="4-cell"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse four-cell-Embryo cDNA library (Long)"  
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMD: 11544199]). The mRNAs were extracted from a pool of 360 embryos at 4-cell stage. Double-stranded cDNAs were synthesized with an oligo(dT) primer (Invitrogen): 5'-PACTAGTCTCTAGTCGCGAGCGCCCTTTT-3' from 10.8ng of mRNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-Linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2Kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 78.3%; Score 18.8; DB 7; Length 132;  
 Best Local Similarity 70.8%; Pred. No. 2.8e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGCTGCGCCGCGTACGCGSAC 24  
 |||||:||||:||||:||||:  
 Db 86 GTCTGCGCGCGCGGAGCGCGCAC 63

## RESULT 3

BB869974/c

LOCUS BB869974 RIKEN full-length enriched, 13 days embryo spinal cord Mus 247 bp mRNA linear EST 27-NOV-2001  
 DEFINITION musculus cDNA clone G630017E12 5', mRNA sequence.

ACCESSION BB869974

BB869974.1 GI:17116184

EST.

Mus musculus (house mouse)

SOURCE

REFERENCE

AUTHORS

## TITLE

Unpublished (2001)

JOURNAL Contact: Yoshihide Hayashizaki

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic

## FEATURES

source

Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Irawa, M., Ohara, E.,  
 Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,  
 Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.  
 and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,  
 Sugahara, Y., and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.  
 e mouse tissues.

## ORIGIN

Query Match 78.3%; Score 18.8; DB 2; Length 247;  
 Best Local Similarity 70.8%; Pred. No. 2.8e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGCTGCGCCGCGTACGCGSAC 24  
 |||||:||||:||||:||||:  
 Db 100 GTCTGCGCGCGCGGAGCGCGCAC 77

## RESULT 4

CK929651/c

LOCUS CK929651 315 bp mRNA linear EST 22-APR-2004  
 DEFINITION p57mgc\_001191 Normalized Magnaporthe grisea cDNA p57mgc\_001191, mRNA sequence.

ACCESSION CK929651

CK929651.1 GI:45417502

EST.

SOURCE

REFERENCE

AUTHORS

Chen, B., Li, Y., Peng, Y., Dong, H. and Li, D.  
 Large-scale identification of ESTs from Magnaporthe grisea by  
 normalized cDNA library sequencing  
 Unpublished (2004)

JOURNAL

COMMENT

Contact: Baozhan Chen, Youzhi Li  
 Laboratory of Subtropical Bioresource Conservation and Utilization  
 Guangxi University, China Agricultural University, Zhejiang  
 University  
 Daxue Road 100#, Nanning, Guangxi, 530004, China  
 Tel: 0086-771-3239566  
 Fax: 0086-771-3237873  
 Email: chenbs@n.gx.cninfo.net, URL: <http://www.estarray.org>

Seq primer: M13 forward primer.  
FEATURES  
Location/Qualifiers  
1. .315  
/organism="Magnaporthe grisea"  
/mol\_type="mRNA"  
/db\_xref="taxon:148305"  
/clone="p5rmgc.001191"  
/issue\_type="Myceillum, conidium, germinating conidium, swelling appressorium, mature appressorium, penetration peg"  
/dev\_stage="Myceillum, conidium, germinating conidium, swelling appressorium, mature appressorium, penetration peg"  
/clone\_1lb="Normalized Magnaporthe grisea cDNA pGEM-T Easy library"  
/note="Vector: pGEM-T Easy"  
ORIGIN  
Query Match 78.3%; Score 18.8; DB 7; Length 315;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCCGTGCGCCGTCGACGCGAC 24  
Db 193 GTCCGTGCGCCGTCGACGCGAC 170  
RESULT 5  
BB868549/c 338 bp mRNA linear EST 27-NOV-2001  
LOCUS BB868549 RIKEN full-length enriched, 16 days neonate male  
DEFINITION diencephalon Mus musculus cDNA clone G63006D18 5', mRNA sequence.  
ACCESSION BB868549  
VERSION BB868549.1 GI:17114759  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 338)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hizamoto,K., Hiroaka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Komuro,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)  
TITLE Unpublished (2001)  
JOURNAL Contact: Yoshihide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Saito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komuro,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagii,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matsuhira,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Komuro,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,

Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
e mouse tissues.  
FEATURES  
Location/Qualifiers  
1. .338  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="G63006D18"  
/sex="male"  
/issue\_type="dienecephalon"  
/dev\_stage="16 days neonate"  
/clone\_1lb="RIKEN full-length enriched, 16 days neonate male dienecephalon"  
ORIGIN  
Query Match 78.3%; Score 18.8; DB 2; Length 338;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCCGTGCGCCGTCGACGCGAC 24  
Db 123 GTCCGTGCGCCGTCGACGCGAC 100  
RESULT 6  
BY280059/c 396 bp mRNA linear EST 11-DEC-2002  
LOCUS BY280059 RIKEN full-length enriched, visual cortex Mus musculus  
DEFINITION cDNA clone K430344N19 5', mRNA sequence.  
ACCESSION BY280059  
VERSION BY280059.1 GI:26470396  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 396)  
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nishida,I., Osato,R., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Wagii,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schombach,C., Gojohori,T., Balderelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Brad,D., Brusic,V., Choitha,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochik,I.V., Lee,Y., Lennard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Nomura,K., Okido,T., Pavan,W.J., Perlee,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Sempile,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Varado,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wymshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hizamoto,K., Kawai,J., Kawai,K., Kawai,X., Kawai,Y., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Aizawa,K., Akakawa,T., Fukuda,S., Hara,A., Hashizume,M., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,B. and Hayashizaki,Y.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
TITLE  
JOURNAL

MEDLINE  
PUBMED  
22354683  
12466851  
COMMENT  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/

Alizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Watanuki, A., Muramatsu, M. and  
Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1575-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Michela Fagioli and Takao K. Hensch ( )  
Laboratory for Neuronal Circuit Development Brain Science Institute  
RIKEN 2-1 Hironawa, Wako-shi, Saitama 351-0198 Japan.) whose  
assistance we gratefully acknowledge. Please visit our web site  
(http://genome.gsc.riken.go.jp) for further details.

FEATURES  
source  
1. .396  
Location/Qualifiers

/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="K430344N19"  
/issue\_type="visual cortex"  
/clone\_lib="RIKEN full-length enriched, visual cortex"

## ORIGIN

Query Match 78.3%; Score 18.8; DB 5; Length 396;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCGTGTGGCCGTSACSGSAC 24  
|||||:|||||:|||||:  
98 GTCCGTGGGCGCGGACGCGAC 75

## RESULT 7

LOCUS CK929166 428 bp mRNA linear EST 22-APR-2004  
DEFINITION p5imgc\_000411 Normalized Magnaporthe grisea cDNA p5imgc\_000411, mRNA sequence.  
ACCESSION CK929166  
VERSION CK929166.1 GI:45417017  
KEYWORDS EST.  
SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)  
ORGANISM Magnaporthe grisea  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetes Incertae sedis; Magnaporthaceae; Magnaporthe.

REFERENCE  
AUTHORS Chen, B., Li, Y., Peng, Y., Dong, H. and Li, D.  
TITLE Large-scale identification of ESTs from Magnaporthe grisea by  
normalized cDNA library sequencing

JOURNAL  
COMMENT

Unpublished (2004)  
Contact: Baoshan Chen, Youzhi Li  
Laboratory of Subtropical Bioresource Conservation and Utilization  
University  
Guangxi University, China Agricultural University, Zhenjiang  
Daxue Road 100#, Nanning, Guangxi, 530004, China  
Tel: 0086-771-3239566  
Fax: 0086-771-3237873  
Email: chenbs@n.yn.cninfo.net, URL: http://www.estarray.org  
Seq primer: M13 forward primer.

## FEATURES

## source

1. .428  
Location/Qualifiers  
/organism="Magnaporthe grisea"  
/mol\_type="mRNA"  
/db\_xref="taxon:148305"  
/clone="p5imgc\_000411"  
/issue\_type="Myceallium, conidium, germinating conidium,  
swelling appressorium, mature appressorium, penetration  
peg"  
/dev\_stage="Myceallium, conidium, germinating conidium,  
swelling appressorium, mature appressorium, penetration  
peg"  
/clone\_lib="Normalized Magnaporthe grisea cDNA p5imgc\_000411  
library"  
/note="Vector: p5imgc-T Easy"

## ORIGIN

Query Match 78.3%; Score 18.8; DB 7; Length 428;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCGTGTGGCCGTSACSGSAC 24  
|||||:|||||:|||||:  
DB 181 GTCCGTGGGCGCGGACGCGAC 158

RESULT 8  
LOCUS BB863614/c

DEFINITION BB863614 RIKEN full-length enriched, pooled cell lines,  
RCB-0544, etc. Mus musculus cDNA clone G430038G07 5', mRNA sequence.  
ACCESSION BB863614  
VERSION BB863614.1 GI:17109824  
KEYWORDS EST.

## SOURCE

## ORGANISM

Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

1 (bases 1 to 432)  
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,  
Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K.,  
Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M.,  
Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,  
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,  
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Tanaka, T., Tomaru, A., Toya, T., Watanuki, A., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.  
2001)

## TITLE

JOURNAL  
COMMENT

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new



```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:468931"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="MDH10B"
/clone_lib="Scares mouse embryo NMME13.5 14.5"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAAGCGGCGGCGGCGGAATTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo. "
```

REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
AUTHORS	1 (bases 1 to 510)			
TITLE	Piao, Y., Dudekula, D. B., Qian, Y., Martin, P. R., Alba, K., Vescevi, A. L. and Ko, M. S. H.			
JOURNAL	Systematic Analyses of NIA Mouse Neural Stem Cell (Undifferentiated) cDNA Library (Long)			
COMMENT	Unpublished (2002)			
FEATURES	Contact: Dawood B. Dudekula Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: cdna@igun.grc.nia.nih.gov Plate: K0972 row: B column: 03 Seq primer: -21M13 Reverse High quality sequence stop: 510 POLY-A-No.			
SOURCE	Location/Qualifiers			
	1. 510			
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/strain="CD1"			
	/db_xref="niaEST:K0972B03-5N"			
	/db_xref="taxon:10090"			
	/clone="NIA:K0972B03 IMAGE:30091118"			
	/dev_stage="Adult"			
	/lab_host="DH10B"			
	/clone_idb="NIA Mouse Neural Stem Cell (Undifferentiated) cDNA Library (Long)"			
	/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH ( <a href="http://igun.grc.nia.nih.gov/cdna">http://igun.grc.nia.nih.gov/cdna</a> ). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Dr. Angelo L. Vescevi (Institute for Stem Cell Research, Italy). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTTCTAGATCCGACGCGCCGCTTTTCTTTT-3'] from 2.0 microgram of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker IL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.8 kb. The library was constructed by Yulan Piao."			
ORIGIN	Query Match 78.3%; Score 18.8; DB 6; Length 510; Best Local Similarity 70.8%; Pred. No. 2.7e+03; Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;			
QY	1 GTCTGATGCGGCGTACGCGSAC 24      :    :    :    : Db 106 GTCTGGGGGCGCGGACGCGCAC 83			
RESULT 12	CB606070/c 516 bp mRNA linear EST 16-MAY-2003			
LOCUS	CB606070			
DEFINITION	AMGNMUC:NRH7-00021-A10-A nrh7 (10850) Rattus norvegicus cDNA			
ACCESSION	CB606070			
VERSION	EST.			
KEYWORDS	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus			

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 516)  
 Angen EST Program.  
 Angen Rat EST Program  
 Unpublished (2003)  
 Contact: Dan Fitzpatrick  
 Angen, Inc  
 One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA  
 Tel: 805 447-4881  
 Plate: 00021 row: a column: 10.  
 Location/Qualifiers  
 1. 516  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /clone="nrhy7-00021-a10"  
 /clone\_1lb="nrhy7 (10850)"  
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; W Rat hypothalamus adult female Wistar rat avg. insert size 2.3 kb fraction 6 and 7"  
 kb fraction 6 and 7"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 6; Length 516;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GTCTGTGCGCCGCGTACGCGSAC 24  
 |||||:||||:||||:||||:  
 Db 74 GTCTGTGCGCGCCGCGGACGCGCAC 51

RESULT 13  
 CB605972/c  
 LOCUS  
 DEFINITION  
 AMGNMNC:NRHVS-00007-B1-A W Rat hypothalamus (10471) Rattus norvegicus cDNA clone nrhy5-00007-b11 5', mRNA sequence.  
 CB605972  
 CB605972.1 GI:29545584  
 EST.  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 526)  
 Angen EST Program.  
 Angen Rat EST Program  
 Unpublished (2003)  
 Contact: Dan Fitzpatrick  
 Angen, Inc  
 One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA  
 Tel: 805 447-4881  
 Plate: 00007 row: b column: 11.  
 Location/Qualifiers  
 1. 526  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /clone="nrhy5-00007-b11"  
 /clone\_1lb="W Rat hypothalamus (10471)"  
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; W Rat hypothalamus adult female Wistar rat avg. insert size 2.3 kb fraction 6 and 7"  
 kb fraction 6 and 7"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 6; Length 526;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GTCTGTGCGCCGCGTACGCGSAC 24  
 |||||:||||:||||:||||:  
 Db 1 GTCTGTGCGCCGCGGACGCGCAC 51

Db 63 GTCTGTGCGCGCCGCGGACGCGCAC 40

RESULT 14  
 CK928924/c  
 LOCUS  
 DEFINITION  
 p5mgc\_000121 Normalized Magnaporthe grisea cDNA pGEM-T Easy library Magnaporthe grisea cDNA clone p5mgc\_000121, mRNA sequence.  
 CK928924  
 CK928924.1 GI:45416775  
 EST.  
 Magnaporthe grisea (anamorph: Pyricularia grisea)  
 Magnaporthe grisea  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes; Incertae sedis; Magnaporthaceae; Magnaporthe.  
 1 (bases 1 to 538)  
 Chen, B., Li, Y., Peng, Y., Dong, H. and Li, D.  
 Large-scale identification of ESTs from Magnaporthe grisea by normalized cDNA library sequencing  
 Unpublished (2004)  
 Contact: Baoshan Chen, Youzhi Li  
 Laboratory of Subtropical Bioresource Conservation and Utilization  
 Guangxi University, China Agricultural University, Zhejiang University  
 Daxue Road 100#, Nanning, Guangxi, 530004, China  
 Tel: 0086-771-3239566  
 Fax: 0086-771-3237873  
 Email: chenbs@n-gx.cninfo.net, URL: http://www.estarray.org  
 Seq primer: M13 forward primer.  
 Location/Qualifiers  
 1. 538  
 /organism="Magnaporthe grisea"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:148305"  
 /clone="p5mgc\_000121"  
 /tissue\_type="mycelium, conidium, germinating conidium, swelling appressorium, mature appressorium, penetration peg"  
 /dev\_stage="mycelium, conidium, germinating conidium, swelling appressorium, mature appressorium, penetration peg"  
 /clone\_1lb="Normalized Magnaporthe grisea cDNA pGEM-T Easy library"  
 /note="Vector: pGEM-T Easy"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 7; Length 538;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GTCTGTGCGCCGCGTACGCGSAC 24  
 |||||:||||:||||:||||:  
 Db 177 GTCTGTGCGCGCCGCGTACGCGCAC 154

RESULT 15  
 BF462782/c  
 LOCUS  
 DEFINITION  
 UI-M-CG0b-bnh-e-07-0-UI-a1 NIH BMAP Ret4.s2 Mus musculus cDNA clone UI-M-CG0b-bnh-e-07-0-UI 3', mRNA sequence.  
 BF462782  
 BF462782.1 GI:11531965  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 551)  
 Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 Normalization and subtraction: two approaches to facilitate gene discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 97044477

PUBMED  
COMMENT 8889548  
Contract: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mes@mail.nih.gov  
Oligo-OT track not found, Not I site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
clones from RESEARCH GENERICS. It should be noted that Bento Soares  
is generating a small number of additional specialized  
non-redundant arrays of BMAP cDNAs whose availability will be  
considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=No.

FEATURES  
source Location/Qualifiers  
1..551  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-CG0P-bnh-e-07-0-UI"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NIH\_BMAP\_Ret4\_S2"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; The  
NIH BMAP Ret4\_S2 library is a subtraced library,  
ultimately derived from mouse retina tissue libraries at  
various stages of development. For a detailed description  
of the library from which this clone was derived, please  
visit our web site at brainest.eng.uiowa.edu. The tissue  
for this library was contributed by Dr. Xin-Yuan Fu, Yale  
University School of Medicine  
TAG\_SEQ=None found"

ORIGIN  
Query Match 78.3%; Score 18.8; DB 2; Length 551;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCTGTTGGCCGTTAGCGAGC 24  
|||||:||||:||||:|||||  
Db 113 GTCTGGGGGCGCGGAGCGGCAC 90

RESULT 16  
AZ417416 577 bp DNA linear GSS 03-OCT-2000  
LOCUS IM0193N06F Mouse 10kb plasmid UGCGIM library Mus musculus genomic  
DEFINITION clone UGCGIM0193N06 F, genomic survey sequence.  
ACCESSION AZ417416  
VERSION AZ417416.1 GI:10541429  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 577)  
REFERENCE  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
Niederhausern,A. and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
JOURNAL Contact: Robert B. Weiss  
COMMENT University of Utah  
Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606

Fax: 801 585 7177  
Email: ddum@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0193 Row: N Column: 06  
Seq primer: CGTGTAAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 577.

FEATURES  
source Location/Qualifiers  
1..577  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGCGIM0193N06"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UGCGIM library"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gi1473214|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

ORIGIN  
Query Match 78.3%; Score 18.8; DB 8; Length 577;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCTGTTGGCCGTTAGCGAGC 24  
|||||:||||:||||:|||||  
Db 411 GTCTGGGGGCGCGGAGCGGCAC 434

RESULT 17  
BB642652/c 610 bp mRNA linear EST 26-OCT-2001  
LOCUS BB642652 RIKEN full-length enriched, adult retina Mus musculus cDNA  
DEFINITION clone A930028E09 5', mRNA sequence.  
ACCESSION BB642652  
VERSION BB642652.1 GI:16477487  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 610)  
REFERENCE  
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,  
Hara,A., Hizumoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,O.,  
Kono,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,  
Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
JOURNAL Contact: Yoshihide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)





prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES  
source

1. 619  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="241004013"  
/cell\_type="ES cells"  
/lab\_host="GDR"  
/clone\_11b="RIKEN full-length enriched, ES cells"  
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'  
GAGAGAGAGATTCGAGTATTAATTAATCCCGCCCCCCC 3'." cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adaptor of sequence 15' GAGAGAGATTCGAGTATTAATTAATCCCGCCCCCCC 3'."

## ORIGIN

Query Match 78.3%; Score 18.8; DB 6; Length 619;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGTGCGCGTACGCGSAC 24  
|||||:||||:||||:||||:||||:  
98 GTCTGTGCGCGCGCGCGCAC 75

Db

## RESULT 19

BU701308 641 bp mRNA linear EST 15-JUL-2003  
LOCUS BU701308/c  
DEFINITION UT-M-EXO-bwy-k-12-0-UT.1 NIH\_BMAP\_EXO Mus musculus cDNA clone  
IMAGE:5704835 5', mRNA sequence.

ACCESSION BU701308  
VERSION BU701308.1 GI:23624943  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE NIH-MGC <http://mgs.nci.nih.gov/>.  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
CDNA library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA library arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:  
<http://image.lnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: PYX-5.

FEATURES  
source  
Location/Qualifiers  
1. 641  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGR:5704835"

## ORIGIN

Query Match 78.3%; Score 18.8; DB 5; Length 641;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGTGCGCGTACGCGSAC 24  
|||||:||||:||||:||||:||||:  
119 GTCTGTGCGCGCGCGCGCAC 96

Db

## RESULT 20

CN721004 672 bp mRNA linear EST 18-MAY-2004  
LOCUS CN721004/c  
DEFINITION E0814B04-5 NIA Mouse four-cell-Embryo cDNA library (long) Mus musculus cDNA clone NIA:E0814B04 IMAGE:30906927 5', mRNA sequence.

ACCESSION CN721004  
VERSION CN721004.1 GI:47490389  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE 1 (bases 1 to 672)  
JOURNAL Sharov,A.A., Piao,Y., Matoba,R., Dudekula,D.B., Qian,Y., Vamburen,V., Falco,G., Martin,P.R., Stegg,C.A., Bassez,U.C., Wang,Y., Carter,M.G., Hamatani,T., Alba,K., Akutsu,H., Sharova,L., Tanaka,T.S., Kimber,W.L., Yoshikawa,T., Jaradat,S.A., Pantano,S., Nagaraja,R., Boheler,K.R., Taub,D., Hodas,R.J., Longo,D.L., Schlessinger,D., Keller,J., Klotz,E., Kelsoe,G., Umezawa,A., Vesceva,A.L., Rossant,J., Hide,W., and Ko,M.S.  
COMMENT Transcriptional analysis of mouse stem cells and early embryos  
Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: [cdna@igsun.grc.nia.nih.gov](mailto:cdna@igsun.grc.nia.nih.gov)  
Plate: E0814 row: B column: 04  
Seq primer: M13 Reverse  
High quality sequence stop: 672  
POLYA=No.

FEATURES  
source

Location/Qualifiers  
1. 672  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="nabEST:E0814B04-5"  
/db\_xref="taxon:10090"  
/clone="NIA:E0814B04 IMAGE:30906927"  
/tissue\_type="4-cell stage embryo"  
/dev\_stage="4-cell"  
/lab\_host="DH10B"

/clone\_lib="NIA Mouse four-cell-Embryo cDNA library (Long)"  
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://1gsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). The mRNAs were extracted from a pool of 360 embryos at 4-cell stage. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTGATCTAGATCGGACGGCCCTTTTCTTTT-3'] from 10.8ng of mRNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker lB-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2Kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 78.3%; Score 18.8; DB 7; Length 672;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 GTCCGTGCGCCGCTACGCGSAC 24  
 |||||:||||:||||:||||:||||:  
 Db 86 GTCCGTGCGCGCCGCGGACGCGCAC 63

RESULT 21 717 bp mRNA linear EST 20-SEP-2001  
 B1732637/c LOCUS  
 DEFINITION 60355381F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:5362485 5', mRNA sequence.

ACCESSION B1732637  
 VERSION B1732637.1 GI:15709650  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 717)  
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strusberg, Ph.D.  
 Email: csgapbs-remail.nih.gov  
 Tissue Procurement: The Cepko Laboratory  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at:  
 http://image.llnl.gov  
 Plate: LLM11922 row: b column: 22  
 High quality sequence stop: 646.  
 Location/Qualifiers

## FEATURES

source  
 1. 717  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5362485"  
 /tissue\_type="retina"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 94"  
 /note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally; Oligo-dT primed.  
 Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 78.3%; Score 18.8; DB 4; Length 717;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 GTCCGTGCGCCGCTACGCGSAC 24  
 |||||:||||:||||:||||:||||:  
 Db 79 GTCCGTGCGCGCCGCGGACGCGCAC 56

RESULT 22 760 bp DNA linear GSS 17-DEC-2002  
 BZ550710 LOCUS  
 DEFINITION pacsl-60.2788.y2 pacsl-60 Pseudomonas aeruginosa genomic clone  
 pacsl-60.2788, genomic survey sequence.

ACCESSION BZ550710  
 VERSION BZ550710.1 GI:27154291  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 760)  
 AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol. (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: ckraymond@u.washington.edu  
 Class: Shotgun.

## FEATURES

source  
 1. 760  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="1-60"  
 /db\_xref="taxon:287"  
 /clone="pacsl-60.2788"  
 /clone\_lib="pacsl-60"  
 /note="clinical isolate 1-60 whole genomic shotgun library."

## ORIGIN

Query Match 78.3%; Score 18.8; DB 8; Length 760;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 GTCCGTGCGCCGCTACGCGSAC 24  
 |||||:||||:||||:||||:||||:  
 Db 450 GTCCGTGCTGCGCGGACGCGCAC 473

RESULT 23 780 bp mRNA linear EST 12-JUN-2000  
 AU080793/c LOCUS  
 DEFINITION AU080793 Sugano mouse brain mchb Mus musculus cDNA clone MNCB-6078  
 5', mRNA sequence.

ACCESSION AU080793  
 VERSION AU080793.1 GI:6085547  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
AUTHORS	1 (bases 1 to 780)
TITLE	Hashimoto, K., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Toyoda, A., Suzuki, Y., Sasaki, M., and Sugano, S.
JOURNAL	Isolation of full-length cDNA clones from a mouse brain cDNA library made by oligo-capping method
COMMENT	Unpublished (1999) Contact: Kazuyuki Hashimoto Division of Genetic Resources National Institute of Infectious Diseases 2-3-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan Email: khashi@nih.go.jp URL: http://www.nih.go.jp/yoken/genbank/. Location/Qualifiers
FEATURES	1..780
SOURCE	/organism="Mus musculus" /mol_type="mRNA" /strain="C57BL" /db_xref="taxon:10090" /clone="MNCh-6078" /sex="female" /dev_stage="adult" /lab_host="TOP10" /clone_lib="Sugano mouse brain mncb" /note="Organ: brain; Vector: pME18S-FU3; 1st strand cDNA was primed with an oligo(dT) primer ATGCGGCCCTTTTTCCTTTTTT"; double-stranded cDNA was ligated to a DraIII adaptor (TTGTGGCTACTGG), digested and cloned into distinct DraIII sites of the pME18S-FU3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Size selection was performed by Sugano et al.(University of Tokyo, Institute of Medical Science). Custom primer used for sequencing: 5' end primer [CTTGCTGCTTAAGCTGC], 3' end primer [CGACCTGCAGCTGACGACA]"
ORIGIN	
Query Match	78.3%; Score 18.8; DB 1; Length 780;
Best Local Similarity	70.8%; Pred. No. 2.7e+03;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0.	
Cy	1 GTCCGSGTGGCCGGSNACSACGSC 24      :     :     :
Db	94 GTCCTGGGCGCGGGGACGCGCAC 71
RESULT 24	
BZ565923/c	
LOCUS	BZ565923 782 bp DNA linear GSS 17-DEC-2002
DEFINITION	pacs2-164_5820_v2 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION	pacs2-164_5820, genomic survey sequence.
VERSION	BZ565923
KEYWORDS	BZ565923.1 GI:27194532
SOURCE	GSS
ORGANISM	Pseudomonas aeruginosa
TAXONOMY	Pseudomonas aeruginosa Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE	1 (bases 1 to 782) Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R., and Olsen, M.V.
AUTHORS	Whole-genome-sequence variation among multiple isolates of Pseudomonas aeruginosa library
TITLE	J. Bacteriol. (2002) In press
JOURNAL	Contact: Chris K. Raymond
COMMENT	Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel.: 2062216954 Fax: 2066857244 Email: cgraymond@u.washington.edu Class: shotgun.

```

FEATURES
source
    Location/Qualifiers
        1..782
            /organism="Pseudomonas aeruginosa"
            /mol_type="genomic DNA"
            /strain="2-164"
            /db_xref="taxon:287"
            /clone="pacs2-164_5820"
            /clone_1lb="pacs2-164"
            /note="clinical isolate 2-164 Whole genomic shotgun library."

ORIGIN
Query Match      78.3%; Score 18.8; DB 8; Length 782;
Best Local Similarity 70.8%; Pred. No. 2.7e+03;
Matches 17, Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTCTGTGGCCGCGTACGCGSAC 24
|||||:||||||:::||
Db       551 GTCCGCCTGGCGCGTGACCCCCAC 528

RESULT 25
AU080514/c      793 bp      mRNA      linear      EST 12-JUL-2000
LOCUS AU080514 Sugano mouse brain mchb Mus musculus cDNA clone MNCD-5748
DEFINITION
5', mRNA sequence.
ACCESSION
AU080514
KEYWORDS
AUI080514.1 GI:6085268
SOURCE
EST.
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 793)
Hashimoto,K., Kusuda,T., Tanuma,R., Ito,A., Hirata,M., Toyoda,A.,
Suzuki,Y., Sasaki,M. and Sugano,S.
Isolation of full-length cDNA clones from a mouse brain cDNA
library made by oligo-capping method
Unpublished (1999)
COMMENT
Contact: Katsuyuki Hashimoto
Division of Genetic Resources
National Institute of Infectious Diseases
23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
Email: khashi@nih.go.jp
URL: http://www.nih.go.jp/yoken/genbank/.
LOCATION/QUALIFIERS
1..793
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="MNCD-5748"
/sex="female"
/dev_stage="adult"
/lab_host="TOP10"
/clone_1lb="Sugano mouse brain mchb"
/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA
was primed with an oligo(dT) primer
ATGGGCCCTTTTYYYTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [GTTGGCCTACTGCG], digested and
cloned into distinct DraIII sites of the pME18S-FL3. XhoI
sites just outside the DraIII sites can be used to isolate
the cDNA insert. Size selection was performed to exclude
fragments <1.5 kb. Library was constructed by Sugano et
al.(University of Tokyo, Institute of Medical Science).
Custom primer used for sequencing: 5' end primer
[CTTCNGCTCTAAAGCTGCG] , 3' end primer
[GCACTGCAGCTGCAGCAC]."

ORIGIN
Query Match      78.3%; Score 18.8; DB 1; Length 793;
Best Local Similarity 70.8%; Pred. No. 2.7e+03;
Matches 17, Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 GTCTGTGGCGCGGTACGCGSAC 24  
 Db 95 GTCTGTGGCGCGCGGTACGCGSAC 72

RESULT 26  
 CNS077KA/c 907 bp DNA linear GSS 08-JUL-2001  
 LOCUS T7 end of clone BB0A012E03 of library BB0A from strain CBS 4732  
 DEFINITION of *Pichia angusta*, genomic survey sequence.  
 ACCESSION AL432800  
 VERSION AL432800.1 GI:12216214  
 KEYWORDS GSS.  
 ORGANISM *Pichia angusta*  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE 1 (bases 1 to 907)  
 Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bojotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Pottier, S., Saurin, W., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
 FEMS Lett. 487 (1), 3-12 (2000)

TITLE  
 JOURNAL FEMS Lett. 487 (1), 76-81 (2000)  
 MEDLINE 20584723  
 PUBMED 11152888  
 REFERENCE 3 (bases 1 to 907)  
 AUTHORS Direct Submission  
 JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seque@genoscope.cns.fr - Web: www.genoscope.cns.fr)  
 COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.  
 Location/Qualifiers  
 1..907  
 /organism="Pichia angusta"  
 /mol\_type="genomic DNA"  
 /strain="CBS 4732"  
 /db\_xref="taxon:4905"  
 /clone="BB0A012E03"  
 /clone\_1fb="BB0AA"  
 /note="end : T7"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 9; Length 907;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGTGGCGCGGTACGCGSAC 24  
 Db 302 GTCTGTGGCGCGGTACGCGSAC 279

RESULT 27  
 B2568254/c 909 bp DNA linear GSS 17-DEC-2002  
 LOCUS pacs2-164\_7441.y2 pacs2-164 *Pseudomonas aeruginosa* genomic clone  
 DEFINITION pacs2-164\_7441, genomic survey sequence.  
 ACCESSION B2568254  
 VERSION B2568254.1 GI:27200910  
 KEYWORDS GSS.  
 SOURCE *Pseudomonas aeruginosa*  
 ORGANISM *Pseudomonas aeruginosa*  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 909)  
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.  
 Whole-Genome-Sequence variation among multiple isolates of *Pseudomonas aeruginosa* library J. Bacteriol. (2002) in press  
 Contact: Chris K. Raymond  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel.: 2062216854  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

FEATURES  
 source  
 1..909  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pacs2-164\_7441"  
 /clone\_1fb="pacs2-164"  
 /note="clinical isolate 2-164 whole genomic shotgun library."

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 8; Length 909;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGTGGCGCGGTACGCGSAC 24  
 Db 348 GTCTGTGGCGCGGTACGCGSAC 325

RESULT 28  
 CNS079PU 927 bp DNA linear GSS 08-JUL-2001  
 LOCUS T3 end of clone BB0A031C05 of library BB0A from strain CBS 4732  
 DEFINITION of *Pichia angusta*, genomic survey sequence.  
 ACCESSION AL435592  
 VERSION AL435592.1 GI:12219005  
 KEYWORDS GSS.  
 SOURCE *Pichia angusta*  
 ORGANISM *Pichia angusta*  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE 1 (bases 1 to 927)  
 Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bojotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Pottier, S., Saurin, W., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
 FEMS Lett. 487 (1), 3-12 (2000)

TITLE  
 JOURNAL FEMS Lett. 487 (1), 3-12 (2000)  
 MEDLINE 20584711  
 PUBMED 11152876  
 REFERENCE 2 (bases 1 to 927)

**AUTHORS** Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F. and Dujon, B.  
**TITLE** Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia angusta*  
**JOURNAL MEDLINE** 20584723  
**PUBMED** 11152888  
**REFERENCE** 3 (bases 1 to 927)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : beqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
**COMMENT** This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvatum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.  
**FEATURES** location/Qualifiers  
 source  
 1..927  
 /organism="Pichia angusta"  
 /mol\_type="genomic DNA"  
 /strain="CBS 4732"  
 /db\_xref="taxon:4905"  
 /clone="BB0AA031C05"  
 /clone\_1lb="BB0AA"  
 /note="end : T3"  
**ORIGIN**  
 Query Match 78.3% Score 18.8; DB 9; Length 927;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 1 GTCTGCTGCGCCSGTSCGCGSAC 24  
 |||||:||||:||||:||||:  
 190 GTCTGCTGCGCGAGACGCGGAC 167  
**RESULT 29** 1024 bp DNA linear GSS 08-JUL-2001  
**LOCUS** CNS0795W  
**DEFINITION** T3 end of clone BB0AA026E07 of library BB0AA from strain CBS 4732 of *Pichia angusta*, genomic survey sequence.  
**ACCESSION** AL434874  
**VERSION** AL434874.1 GI:12218287  
**KEYWORDS** GSS.  
**SOURCE** *Pichia angusta*  
**ORGANISM** *Pichia angusta*  
**REFERENCE** Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; *Pichia*.  
**AUTHORS** 1 (bases 1 to 1024)  
 Soucier, J. L., Aigle, M., Artiguenave, F., Blandin, G., Bolicin-Fukuhara, M., Bon, E., Broctier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Leplingle, A., Llorente, B., Malpertuy, A., Neveglise, C., Olier-Kalogeropoulos, O., Potier, S., Sautin, W., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M., Wincker, P. and Weissensbach, J.  
**TITLE** Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
**JOURNAL MEDLINE** 20584711  
**PUBMED** 11152876  
**REFERENCE** 2 (bases 1 to 1024)  
**AUTHORS** Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F. and Dujon, B.  
**TITLE** Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia angusta*  
**JOURNAL MEDLINE** 20584723  
**PUBMED** 11152888  
**REFERENCE** 3 (bases 1 to 1024)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : beqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
**COMMENT** This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvatum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.  
**FEATURES** location/Qualifiers  
 source  
 1..1024  
 /organism="Pichia angusta"  
 /mol\_type="genomic DNA"  
 /strain="CBS 4732"  
 /db\_xref="taxon:4905"  
 /clone="BB0AA026E07"  
 /clone\_1lb="BB0AA"  
 /note="end : T3"  
**ORIGIN**  
 Query Match 78.3% Score 18.8; DB 9; Length 1024;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 1 GTCTGCTGCGCCSGTSCGCGSAC 24  
 |||||:||||:||||:||||:  
 65 GTCTGCTGCGCGAGACGCGGAC 88  
**RESULT 30** 1325 bp mRNA linear HTC 03-APR-2004  
**LOCUS** AK010405/c  
**DEFINITION** Mus musculus ES cells cDNA, RIKEN full-length enriched library, clone:2410004013 product:porcupine homolog (Drosophila), full insert sequence.  
**ACCESSION** AK010405  
**VERSION** AK010405.1 GI:12845824  
**KEYWORDS** HTC; CAP trapper.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**AUTHORS** 1  
 Carninci, P. and Hayashizaki, Y.  
**TITLE** High-efficiency full-length cDNA cloning  
**JOURNAL MEDLINE** 99279253  
**PUBMED** 10349636  
**REFERENCE** 2  
 Carninci, P. and Hayashizaki, Y.  
**AUTHORS** Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
**TITLE** Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
**JOURNAL MEDLINE** 2049374  
**PUBMED** 11042159  
**REFERENCE** 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, U., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, B., Matsubiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.



**THIS PAGE IS BLANK**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 13:19:40 ; Search time 394.857 Seconds  
(without alignment)  
359.811 Million cell updates/sec

Title: US-10-017-471B-2

Perfect score: 24

Sequence: 1 gtccgtgctgcgcgtacscgcsac 24

Scoring table: IDENTITY NUC

Gapop 10'-0', Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : N\_Geneseq\_16Dec04:\*

1: Geneseq19808:\*

2: Geneseq19908:\*

3: Geneseq20008:\*

4: Geneseq20018:\*

5: Geneseq20028:\*

6: Geneseq20038:\*

7: Geneseq20048:\*

8: Geneseq20058:\*

9: Geneseq20068:\*

10: Geneseq20078:\*

11: Geneseq20088:\*

12: Geneseq20098:\*

13: Geneseq20108:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	91.7	24	6	ABN86367
2	22	91.7	24	10	ADH14796
3	19.2	80.0	4346	6	ABN86379
4	19.2	80.0	4346	10	ADH14796
5	18.8	78.3	669	11	ABD11456
6	18.8	78.3	810	11	ABD11456
7	18.8	78.3	1100	13	ADH64127
8	18.8	78.3	1136	13	ADH64127
9	18.8	78.3	1136	13	ADH64127
10	18.8	78.3	1136	13	ADH64127
11	18.8	78.3	1136	13	ADH64127
12	18.8	78.3	1136	13	ADH64127
13	18.8	78.3	1136	13	ADH64127
14	18.8	78.3	1136	13	ADH64127
15	18.8	78.3	1136	13	ADH64127
16	18.8	78.3	1136	13	ADH64127
17	18.8	78.3	1136	13	ADH64127
18	18.8	78.3	1136	13	ADH64127
19	18.8	78.3	1136	13	ADH64127
20	18.8	78.3	1136	13	ADH64127

21	17.8	74.2	1455	12	ADP04866	ADP04866 Sea equit
22	17.4	72.5	492	11	ABD07378	ABD07378 Pseudomon
23	17.4	72.5	654	11	ABD01735	ABD01735 Pseudomon
24	17.4	72.5	744	4	AA554125	AA554125 Pseudomon
25	17.4	72.5	744	8	AA542298	AA542298 Pseudomon
26	17.4	72.5	759	11	ABD01726	ABD01726 Pseudomon
27	17.4	72.5	843	11	ACH97630	ACH97630 Klebsiell
28	17.4	72.5	1026	11	ABD01715	ABD01715 Pseudomon
29	17.4	72.5	1212	6	ABQ90070	ABQ90070 M. capsul
30	17.4	72.5	1578	11	ABD07361	ABD07361 Pseudomon
31	17.4	72.5	1827	11	ABD12449	ABD12449 Pseudomon
32	17.4	72.5	1929	11	ABD12597	ABD12597 Pseudomon
33	17.4	72.5	1962	11	ABD12378	ABD12378 Pseudomon
34	17.4	72.5	1998	12	AAQ11127	AAQ11127 Sequence
35	17.4	72.5	2351	12	ADP63621	ADP63621 Novel hum
36	17.4	72.5	2469	11	ABD07304	ABD07304 Pseudomon
37	17.4	72.5	2553	4	AA554053	AA554053 Pseudomon
38	17.4	72.5	2553	8	ACA42124	ACA42124 Pseudomon
39	17.4	72.5	3135	11	ABD07318	ABD07318 Pseudomon
40	17.2	71.7	585	11	ABD07403	ABD07403 Pseudomon
41	17.2	71.7	636	11	ADP65673	ADP65673 Human EST
42	17.2	71.7	672	5	AA577349	AA577349 DNA encod
43	17.2	71.7	825	4	ABL16533	ABL16533 Drosophil
44	17.2	71.7	951	6	ABQ90286	ABQ90286 M. capsul
45	17.2	71.7	951	6	ABQ90429	ABQ90429 M. capsul
46	17.2	71.7	963	8	ACA26815	ACA26815 Pseudomon
47	17.2	71.7	978	11	ABD07780	ABD07780 Pseudomon
48	17.2	71.7	1146	6	AB553735	AB553735 Aspergill
49	17.2	71.7	1422	8	ACA26887	ACA26887 Pseudomon
50	17.2	71.7	1482	11	ABD11085	ABD11085 Pseudomon
51	17.2	71.7	1614	11	ABD11003	ABD11003 Pseudomon
52	17.2	71.7	1623	11	ABD11046	ABD11046 Pseudomon
53	17.2	71.7	1632	8	ADA71068	ADA71068 Rice gene
54	17.2	71.7	1667	12	AD143097	AD143097 Plant tra
55	17.2	71.7	1821	13	ADT48652	ADT48652 Bacterial
56	17.2	71.7	1866	8	ACA37690	ACA37690 Pseudomon
57	17.2	71.7	2070	13	ADP01214	ADP01214 Pseudomon
58	17.2	71.7	2127	4	ABL11312	ABL11312 Drosophil
59	17.2	71.7	2189	11	ADP02770	ADP02770 Human CDN
60	17.2	71.7	2688	13	ADP01216	ADP01216 Pseudomon
61	17.2	71.7	2712	4	ADP07034	ADP07034 Pseudomon
62	17.2	71.7	2825	4	ABL16532	ABL16532 Drosophil
63	17.2	71.7	2888	3	AAA11239	AAA11239 C. reinha
64	17.2	71.7	3387	5	AA591234	AA591234 DNA encod
65	17.2	71.7	3450	12	ADP06832	ADP06832 Streptomy
66	17.2	71.7	4133	10	ADP15449	ADP15449 S. violac
67	17.2	71.7	4142	12	ADP059812	ADP059812 Novel hum
68	17.2	71.7	9222	8	AA161171	AA161171 Actinosyn
69	17.2	71.7	13268	6	ABV94056	ABV94056 Breast ca
70	17.2	71.7	13268	12	ADN04624	ADN04624 Antipepti
71	17.2	71.7	13268	12	ADP083790	ADP083790 Human tum
72	17.2	71.7	13857	4	AA158199	AA158199 Human pol
73	17.2	71.7	13857	5	ADQ98405	ADQ98405 DNA encod
74	17.2	71.7	13857	9	ADP48165	ADP48165 Novel hum
75	17.2	71.7	36602	13	ADP01210	ADP01210 Pseudomon
76	17.2	71.7	82746	8	AA161224	AA161224 Actinosyn
77	17.2	71.7	88421	6	AA140781	AA140781 88421nt g
78	17.2	71.7	100267	6	ABT11032	ABT11032 Human bre
79	17.2	71.7	103599	4	ABX04971	ABX04971 S. cinna
80	17.2	71.7	109519	5	AA508693	AA508693 Microtomo
81	17.2	71.7	135638	10	ABX34289	ABX34289 S. atrocol
82	17.2	71.7	1656	8	ACA38211	ACA38211 Pseudomon
83	16.8	70.0	300	2	AA214828	AA214828 Human gen
84	16.8	70.0	354	12	ADP28244	ADP28244 Human sec
85	16.8	70.0	378	6	ABN21346	ABN21346 Human ORF
86	16.8	70.0	417	4	AAH99775	AAH99775 Human pro
87	16.8	70.0	429	4	AA522695	AA522695 Human CDN
88	16.8	70.0	873	13	AA522459	AA522459 Bacterial
89	16.8	70.0	884	4	AA522459	AA522459 Human CDN
90	16.8	70.0	1203	4	AA551564	AA551564 Pseudomon
91	16.8	70.0	1203	8	ACA19537	ACA19537 Pseudomon
92	16.8	70.0	1215	11	ABD17838	ABD17838 Pseudomon
93	16.8	70.0	1215	11	ABD17838	ABD17838 Pseudomon

C 94	16.8	70.0	1230	6	ABQ79850	Abq79850 Human TAN	C 167	16.8	70.0	2037	8	ACA68485	ACA68485 Novel hum
C 95	16.8	70.0	1230	12	AD133308	Ad133308 Partial h	C 168	16.8	70.0	2037	8	ABX98950	ABX98950 Novel hum
C 96	16.8	70.0	1254	13	ADT42027	Adt42027 Bacterial	C 169	16.8	70.0	2037	8	ACC81427	ACC81427 Human sec
C 97	16.8	70.0	1263	3	AAZ51245	Aaz51245 Human TAN	C 170	16.8	70.0	2037	8	ACA95751	ACA95751 Novel hum
C 98	16.8	70.0	1263	6	ABQ79849	Abq79849 Human TAN	C 171	16.8	70.0	2037	8	ACD04669	ACD04669 Novel hum
C 99	16.8	70.0	1263	12	AD133307	Ad133307 Partial h	C 172	16.8	70.0	2037	8	ACC88110	ACC88110 Human sec
C 100	16.8	70.0	1263	13	ADSA5825	Adsa5825 Bacterial	C 173	16.8	70.0	2037	8	ACF12772	ACF12772 Human sec
C 101	16.8	70.0	1476	11	ABD16583	Abd16583 Pseudomon	C 174	16.8	70.0	2037	8	ACA96487	ACA96487 Human PRO
C 102	16.8	70.0	1707	8	ADA69639	Ada69639 Rice gene	C 175	16.8	70.0	2037	8	ACA65261	ACA65261 Human PRO
C 103	16.8	70.0	1797	11	ABD15840	Abd15840 Pseudomon	C 176	16.8	70.0	2037	8	ACA73987	ACA73987 Human sec
C 104	16.8	70.0	1887	11	ABD17552	Abd17552 Pseudomon	C 177	16.8	70.0	2037	8	ACA74399	ACA74399 Novel hum
C 105	16.8	70.0	2025	11	ABD16325	Abd16325 Pseudomon	C 178	16.8	70.0	2037	8	ACA96794	ACA96794 Human PRO
C 106	16.8	70.0	2037	4	AA546220	Aa546220 Human DNA	C 179	16.8	70.0	2037	8	ACD10900	ACD10900 CDNA enco
C 107	16.8	70.0	2037	8	ACA89670	ACA89670 CDNA enco	C 180	16.8	70.0	2037	8	ACC91596	ACC91596 Human sec
C 108	16.8	70.0	2037	8	ACA73680	ACA73680 Human sec	C 181	16.8	70.0	2037	8	ACD02931	ACD02931 CDNA enco
C 109	16.8	70.0	2037	8	ACA05995	ACA05995 Human sec	C 182	16.8	70.0	2037	8	ACC87496	ACC87496 Human sec
C 110	16.8	70.0	2037	8	ACA66829	ACA66829 CDNA enco	C 183	16.8	70.0	2037	8	ACC86080	ACC86080 Human sec
C 111	16.8	70.0	2037	8	ACF20404	Acf20404 Human sec	C 184	16.8	70.0	2037	8	ACA65568	ACA65568 Human PRO
C 112	16.8	70.0	2037	8	ACF19790	Acf19790 Human sec	C 185	16.8	70.0	2037	8	ACA94385	ACA94385 Human PRO
C 113	16.8	70.0	2037	8	ACF12078	Acf12078 Human sec	C 186	16.8	70.0	2037	8	ACA98129	ACA98129 Human PRO
C 114	16.8	70.0	2037	8	ACF13243	Acf13243 Human sec	C 187	16.8	70.0	2037	8	ACA91631	ACA91631 Novel hum
C 115	16.8	70.0	2037	8	ACD25346	Acdd25346 Human sec	C 188	16.8	70.0	2037	8	ACA90845	ACA90845 Novel hum
C 116	16.8	70.0	2037	8	ACF00395	Acf00395 Human sec	C 189	16.8	70.0	2037	8	ACD16392	ACD16392 Human sec
C 117	16.8	70.0	2037	8	ACA72452	AcA72452 Novel hum	C 190	16.8	70.0	2037	8	ACD17553	ACD17553 Human sec
C 118	16.8	70.0	2037	8	ACD04976	Acdd04976 Novel hum	C 191	16.8	70.0	2037	8	ACC92210	ACC92210 Human sec
C 119	16.8	70.0	2037	8	ACD18437	Acdd18437 Human sec	C 192	16.8	70.0	2037	8	ACA75067	ACA75067 CDNA enco
C 120	16.8	70.0	2037	8	ACD08444	Acdd08444 Human sec	C 193	16.8	70.0	2037	8	ACA91938	ACA91938 Human PRO
C 121	16.8	70.0	2037	8	ACA88878	ACA88878 Novel hum	C 194	16.8	70.0	2037	8	ACA71582	ACA71582 Human sec
C 122	16.8	70.0	2037	8	ACA70320	ACA70320 Human sec	C 195	16.8	70.0	2037	8	ACC90982	ACC90982 Human sec
C 123	16.8	70.0	2037	8	ACD12542	Acdd12542 Novel hum	C 196	16.8	70.0	2037	8	ACA65992	ACA65992 CDNA enco
C 124	16.8	70.0	2037	8	ACC74457	Acc74457 Human sec	C 197	16.8	70.0	2037	8	ACA95137	ACA95137 CDNA enco
C 125	16.8	70.0	2037	8	ACD16085	Acdd16085 Human sec	C 198	16.8	70.0	2037	8	ACD16699	ACD16699 Human sec
C 126	16.8	70.0	2037	8	ACD25653	Acdd25653 Novel hum	C 199	16.8	70.0	2037	8	ACD15778	ACD15778 Human sec
C 127	16.8	70.0	2037	8	ACD18130	Acdd18130 Human sec	C 200	16.8	70.0	2037	8	ABX16881	ABX16881 Human CDN
C 128	16.8	70.0	2037	8	ACC88417	Acc88417 Human sec	C 201	16.8	70.0	2037	8	ACA97822	ACA97822 Human PRO
C 129	16.8	70.0	2037	8	ACD11771	Acdd11771 Human sec	C 202	16.8	70.0	2037	9	ACA99271	ACA99271 Novel hum
C 130	16.8	70.0	2037	8	ACD18838	Acdd18838 Human sec	C 203	16.8	70.0	2037	9	ACC91903	ACC91903 Human sec
C 131	16.8	70.0	2037	8	ABX98448	Abx98448 Human CDN	C 204	16.8	70.0	2037	9	ACD11314	ACD11314 Novel hum
C 132	16.8	70.0	2037	8	ACD14199	Acdd14199 Human PRO	C 205	16.8	70.0	2037	9	ACD15164	ACD15164 Human sec
C 133	16.8	70.0	2037	8	ACD09979	Acdd09979 Human sec	C 206	16.8	70.0	2037	9	ACD11928	ACD11928 Human sec
C 134	16.8	70.0	2037	8	ACC88724	Acc88724 Human sec	C 207	16.8	70.0	2037	9	ACC96057	ACC96057 Human sec
C 135	16.8	70.0	2037	8	ABX21464	Abx21464 Human sec	C 208	16.8	70.0	2037	9	ACF16620	ACF16620 Human sec
C 136	16.8	70.0	2037	8	ABX75836	Abx75836 Human CDN	C 209	16.8	70.0	2037	9	ACF02738	ACF02738 Human sec
C 137	16.8	70.0	2037	8	ABX98039	Abx98039 Human PRO	C 210	16.8	70.0	2037	9	ACF03045	ACF03045 Human sec
C 138	16.8	70.0	2037	8	ACA57978	ACA57978 Novel hum	C 211	16.8	70.0	2037	9	ACF21632	ACF21632 Human sec
C 139	16.8	70.0	2037	8	ACA57978	ACA57978 Novel hum	C 212	16.8	70.0	2037	9	ACF10316	ACF10316 Human sec
C 140	16.8	70.0	2037	8	ACD14506	Acdd14506 Human PRO	C 213	16.8	70.0	2037	9	ACF78209	ACF78209 Human sec
C 141	16.8	70.0	2037	8	ACC91289	Acc91289 Human sec	C 214	16.8	70.0	2037	9	ACD46914	ACD46914 Human sec
C 142	16.8	70.0	2037	8	ACC89031	Acc89031 Human sec	C 215	16.8	70.0	2037	9	ACD49677	ACD49677 Human sec
C 143	16.8	70.0	2037	8	ACD07228	Acdd07228 Human PRO	C 216	16.8	70.0	2037	9	ACF28444	ACF28444 Human sec
C 144	16.8	70.0	2037	8	ACA67679	ACA67679 Human PRO	C 217	16.8	70.0	2037	9	ACD89134	ACD89134 Human sec
C 145	16.8	70.0	2037	8	ACC81734	Acc81734 Human sec	C 218	16.8	70.0	2037	9	ACD84529	ACD84529 Human PRO
C 146	16.8	70.0	2037	8	ACC89338	Acc89338 Human sec	C 219	16.8	70.0	2037	9	ACD99303	ACD99303 CDNA enco
C 147	16.8	70.0	2037	8	ACC86694	Acc86694 Human sec	C 220	16.8	70.0	2037	9	ADA78343	ADA78343 Human sec
C 148	16.8	70.0	2037	8	ACC89992	Acc89992 Human sec	C 221	16.8	70.0	2037	9	ACF49045	ACF49045 Human sec
C 149	16.8	70.0	2037	8	ACC93131	Acc93131 Human sec	C 222	16.8	70.0	2037	9	ACD09365	ACD09365 Human sec
C 150	16.8	70.0	2037	8	ACA72759	ACA72759 Human PRO	C 223	16.8	70.0	2037	9	ACF12158	ACF12158 Human sec
C 151	16.8	70.0	2037	8	ACA89277	ACA89277 Human sec	C 224	16.8	70.0	2037	9	ACF41392	ACF41392 Human sec
C 152	16.8	70.0	2037	8	ACA70013	ACA70013 Human sec	C 225	16.8	70.0	2037	9	ACF16006	ACF16006 Human sec
C 153	16.8	70.0	2037	8	ACA97156	ACA97156 Novel hum	C 226	16.8	70.0	2037	9	ACF16313	ACF16313 Human sec
C 154	16.8	70.0	2037	8	ACA91152	ACA91152 Novel hum	C 227	16.8	70.0	2037	9	ACD32140	ACD32140 Human sec
C 155	16.8	70.0	2037	8	ACA70934	ACA70934 Human sec	C 228	16.8	70.0	2037	9	ACF18948	ACF18948 Human sec
C 156	16.8	70.0	2037	8	ACA95444	ACA95444 Novel hum	C 229	16.8	70.0	2037	9	ACF09395	ACF09395 Human sec
C 157	16.8	70.0	2037	8	ACC86387	Acc86387 Human sec	C 230	16.8	70.0	2037	9	ACF78516	ACF78516 Human sec
C 158	16.8	70.0	2037	8	ACC90259	Acc90259 Human sec	C 231	16.8	70.0	2037	9	ACF52115	ACF52115 Human sec
C 159	16.8	70.0	2037	8	ACD12867	Acdd12867 Human sec	C 232	16.8	70.0	2037	9	ACF26602	ACF26602 Human sec
C 160	16.8	70.0	2037	8	ACF20097	Acf20097 Human sec	C 233	16.8	70.0	2037	9	ACF24395	ACF24395 Human sec
C 161	16.8	70.0	2037	8	ABX77041	Abx77041 Human PRO	C 234	16.8	70.0	2037	9	ACF63706	ACF63706 Human sec
C 162	16.8	70.0	2037	8	ACA73373	ACA73373 Novel hum	C 235	16.8	70.0	2037	9	ACF50580	ACF50580 Human sec
C 163	16.8	70.0	2037	8	ACA68916	ACA68916 Novel hum	C 236	16.8	70.0	2037	9	ACH08051	ACH08051 Human sec
C 164	16.8	70.0	2037	8	ACA74760	ACA74760 CDNA enco	C 237	16.8	70.0	2037	9	ACF13857	ACF13857 Human sec
C 165	16.8	70.0	2037	8	ACA70627	ACA70627 Human sec	C 238	16.8	70.0	2037	9	ACD41783	ACD41783 Human sec
C 166	16.8	70.0	2037	8	ACD14813	Acdd14813 Human PRO	C 239	16.8	70.0	2037	9	ACF32196	ACF32196 Human sec

C 240	16.8	70.0	2037	9	ACF23474	Act23474 Human sec
C 241	16.8	70.0	2037	9	ACF40164	Act40164 Human sec
C 242	16.8	70.0	2037	9	ACD45686	Act45686 Human sec
C 243	16.8	70.0	2037	9	ACF53343	Act53343 Human sec
C 244	16.8	70.0	2037	9	ACF27523	Act27523 Human sec
C 245	16.8	70.0	2037	9	ACF45361	Act45361 Human sec
C 246	16.8	70.0	2037	9	ACF29979	Act29979 Human sec
C 247	16.8	70.0	2037	9	ACD90055	Act90055 Human sec
C 248	16.8	70.0	2037	9	ACD84836	Act84836 Human PRO
C 249	16.8	70.0	2037	9	ACD98996	Act98996 CDNA enco
C 250	16.8	70.0	2037	9	ACF77288	Act77288 Human sec
C 251	16.8	70.0	2037	9	ACF76981	Act76981 Human sec
C 252	16.8	70.0	2037	9	ACF49966	Act49966 Human sec
C 253	16.8	70.0	2037	9	ACF50273	Act50273 Human sec
C 254	16.8	70.0	2037	9	ACD09672	Act09672 Human sec
C 255	16.8	70.0	2037	9	ACD08751	Act08751 Human sec
C 256	16.8	70.0	2037	9	ACF12465	Act12465 Human sec
C 257	16.8	70.0	2037	9	ACG94973	Act94973 Human sec
C 258	16.8	70.0	2037	9	ACD22692	Act22692 Human sec
C 259	16.8	70.0	2037	9	ACF15392	Act15392 Human sec
C 260	16.8	70.0	2037	9	ACG97487	Act97487 Human sec
C 261	16.8	70.0	2037	9	ACG92517	Act92517 Human sec
C 262	16.8	70.0	2037	9	ACF14164	Act14164 Human sec
C 263	16.8	70.0	2037	9	ACF14471	Act14471 Human sec
C 264	16.8	70.0	2037	9	ACF09702	Act09702 Human sec
C 265	16.8	70.0	2037	9	ACD45993	Act45993 Human sec
C 266	16.8	70.0	2037	9	ACD48142	Act48142 Human sec
C 267	16.8	70.0	2037	9	ACD67873	Act67873 CDNA enco
C 268	16.8	70.0	2037	9	ACF25681	Act25681 Human sec
C 269	16.8	70.0	2037	9	ACF29565	Act29565 Human sec
C 270	16.8	70.0	2037	9	ACD85143	Act85143 Human sec
C 271	16.8	70.0	2037	9	ACD84222	Act84222 Human PRO
C 272	16.8	70.0	2037	9	ACD88213	Act88213 Human sec
C 273	16.8	70.0	2037	9	ACF30900	Act30900 Human sec
C 274	16.8	70.0	2037	9	ACF32503	Act32503 Human sec
C 275	16.8	70.0	2037	9	ACH12163	Act12163 CDNA enco
C 276	16.8	70.0	2037	9	ACH12470	Act12470 CDNA enco
C 277	16.8	70.0	2037	9	ACD40862	Act40862 Human sec
C 278	16.8	70.0	2037	9	ACF18334	Act18334 Human sec
C 279	16.8	70.0	2037	9	ACF08781	Act08781 Human sec
C 280	16.8	70.0	2037	9	ACF31582	Act31582 Human sec
C 281	16.8	70.0	2037	9	ACF52422	Act52422 Human sec
C 282	16.8	70.0	2037	9	ACD50291	Act50291 Human sec
C 283	16.8	70.0	2037	9	ACF38994	Act38994 Human sec
C 284	16.8	70.0	2037	9	ACF26909	Act26909 Human sec
C 285	16.8	70.0	2037	9	ACF25009	Act25009 Human sec
C 286	16.8	70.0	2037	9	ACF46589	Act46589 Human sec
C 287	16.8	70.0	2037	9	ACF28137	Act28137 Human sec
C 288	16.8	70.0	2037	9	ACD89441	Act89441 Human sec
C 289	16.8	70.0	2037	9	ACF64013	Act64013 Human sec
C 290	16.8	70.0	2037	9	ACF60653	Act60653 Human sec
C 291	16.8	70.0	2037	9	ACH12777	Act12777 CDNA enco
C 292	16.8	70.0	2037	9	ACH10200	Act10200 Human sec
C 293	16.8	70.0	2037	9	ACD04055	Act04055 Human sec
C 294	16.8	70.0	2037	9	ACD10593	Act10593 Human sec
C 295	16.8	70.0	2037	9	ACF42620	Act42620 Human sec
C 296	16.8	70.0	2037	9	ACF18641	Act18641 Human sec
C 297	16.8	70.0	2037	9	ACF02431	Act02431 Human sec
C 298	16.8	70.0	2037	9	ACF21939	Act21939 Human sec
C 299	16.8	70.0	2037	9	ACF10623	Act10623 Human sec
C 300	16.8	70.0	2037	9	ACF34075	Act34075 Human sec

## ALIGNMENTS

RESULT 1  
ABN86367 standard; DNA; 24 BP.

XX AC ABN86367;  
XX DT 21-OCT-2002 (first entry)

XX DE S. coelicolor scdA gene internal segment amplifying primer 2.  
XX Antibiotic; bacterium; scdA; afcA; scdR; arpA; barA; actinorhodine; Act;  
KM undecylprodigiosin; Red; PCR; primer; ss.  
XX OS Streptomyces coelicolor.  
XX CA2322241-A1.  
XX PD 23-APR-2002.  
XX PF 23-OCT-2000; 2000CA-02322241.  
XX PR 23-OCT-2000; 2000CA-02322241.  
XX (PLAN-) PLANT BIOSCIENCE LTD.  
XX PA Takano E, Bibb M;  
XX WPI; 2002-501089/54.  
XX DR  
XX PT Modifying antibiotic-producing Streptomyces, to increase, or alter timing  
XX of, antibiotic production, by deleting the scdA or scdR genes.  
XX PS Claim 19; Page 53; 64pp; English.  
XX CC The invention provides a method for modifying an antibiotic-producing  
XX strain of Streptomyces to increase production of antibiotics or to alter  
XX the timing of antibiotic production. The modification is functional  
XX deletion of the scdA gene of S. coelicolor, or its homologues, but is not  
XX deletion of the afcA gene of S. griseus, or the modification is  
XX functional deletion of the scdR gene of S. coelicolor, or its homologues,  
XX but is not deletion of any of S. griseus nor any of S. virginitiae. The  
XX method is particularly used for production of the antibiotics  
XX actinorhodine (Act) and undecylprodigiosin (Red). The present sequence  
XX represents a PCR primer for amplifying an internal segment of the scdA  
XX gene from S. coelicolor M145 total DNA  
XX  
SQ Sequence 24 BP; 2 A; 7 C; 6 G; 4 T; 0 U; 5 Other;  
Query Match 91.7%; Score 22; DB 6; Length 24;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCTGCTGCGCGCTGACGCGAC 24  
DB 1 GTCTGCTGCGCGCTGACGCGAC 24  
RESULT 2  
ID ADE14796 standard; DNA; 24 BP.  
XX AC ADE14796;  
XX DT 29-JAN-2004 (first entry)  
XX DE PCR primer 2 related to Streptomyces coelicolor antibiotic production.  
XX KW antibiotic-producing strain; antibiotic production; scdA gene; scdR gene;  
XX actinorhodin; undecylprodigiosin; PCR; primer; ss.  
XX OS Synthetic.  
XX Streptomyces coelicolor.  
XX US2003124644-A1.  
XX PD 03-JUL-2003.  
XX PF 23-OCT-2001; 2001US-00017471.  
XX PR 23-OCT-2000; 2000US-0242561P.

```
XX (TAKA/) TAKANO E.  
PA (BIBB/) BIBB M J.  
XX  
XX Takano E, Bibb MJ;  
DR WPI, 2003-810983/76.  
XX  
XX Modifying an antibiotic-producing strain of Streptomyces coelicolor or  
PT Streptomyces lividans to increase or alter the timing of antibiotic  
PT production in the strain, comprises functionally deleting in the strain  
PT the scbA or ScbR gene.  
XX  
XX Example 8; Page 9; 33pp; English.  
XX  
XX This invention relates to the novel modification of an antibiotic-  
CC producing strain of Streptomyces coelicolor or Streptomyces lividans to  
CC increase or to alter the timing of antibiotic production in the strain.  
CC The method comprises functionally deleting in the strain the scbA or ScbR  
CC gene. The method is useful in increasing and altering the timing of  
CC antibiotic production (especially actinorhodin and undecylprodigiosin) in  
CC Streptomyces species, particularly Streptomyces coelicolor or  
CC Streptomyces lividans. The present sequence is that of a degenerate PCR  
CC primer which was used for amplification of a region of the Streptomyces  
CC coelicolor ScbA gene during the exemplification of the invention.  
XX  
SQ Sequence 24 BP; 2 A; 7 C; 6 G; 4 T; 0 U; 5 Other;  
  
Query Match 91.7%; Score 22; DB 10; Length 24;  
Best Local Similarity 79.2%; Pred. No. 32;  
Matches 19; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GTCTGTGCGCGTSGACSGSAC 24  
Db 1 GTCTGTGCGCGTSGACSGSAC 24  
  
RESULT 3  
ABN86379  
ID ABN86379 standard; DNA; 4346 BP.  
XX  
XX ABN86379;  
XX  
XX 21-OCT-2002 (first entry)  
XX  
XX ScbA, ScbR and ScbB encoding genes containing DNA sequence.  
XX  
XX Antibiotic; bacterium; scbA; afsA; scbR; arpa; barA; actinorhodine; Act;  
KW undecylprodigiosin; Red; gene; ds.  
XX  
XX Streptomyces coelicolor.  
XX  
XX CA2322241-A1.  
XX  
XX 23-APR-2002.  
XX  
XX 23-OCT-2000; 2000CA-02322241.  
XX  
XX 23-OCT-2000; 2000CA-02322241.  
XX  
XX (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
XX Takano E, Bibb M;  
XX  
XX WPI, 2002-501089/54.  
DR P-PSDB; ABB80940, ABB80941, ABB80942.  
XX  
XX Modifying antibiotic-producing Streptomyces, to increase, or alter timing  
PT of, antibiotic production, by deleting the scbA or scbR genes.  
XX  
XX Claim 19, 20; Fig 14; 64pp; English.  
XX  
XX The invention provides a method for modifying an antibiotic-producing
```

```
CC strain of Streptomyces to increase production of antibiotics or to alter  
CC the timing of antibiotic production. The modification is functional  
CC deletion of the scbA gene of S. coelicolor, or its homologues, but is not  
CC functional deletion of the afsA gene of S. griseus, or the modification is  
CC functional deletion of the scbR gene of S. coelicolor, or its homologues,  
CC but is not deletion of arpa of S. griseus nor barA of S. virgatae. The  
CC method is particularly used for production of the antibiotics  
CC actinorhodine (Act) and undecylprodigiosin (Red). The present sequence  
CC represents the nucleic acid sequence of the region containing the S.  
CC coelicolor ScbA, ScbR and ScbB proteins  
XX  
SQ Sequence 4346 BP; 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;  
  
Query Match 80.0%; Score 19.2; DB 6; Length 4346;  
Best Local Similarity 75.0%; Pred. No. 3,7e+02;  
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 GTCTGTGCGCGTSGACSGSAC 24  
Db 1250 GTCTGTGCGCGTSGACSGSAC 1273  
  
RESULT 4  
ADE14794  
ID ADE14794 standard; DNA; 4346 BP.  
XX  
XX ADE14794;  
XX  
XX 29-JAN-2004 (first entry)  
XX  
XX Streptomyces coelicolor partial genome sequence 2.  
XX  
XX Streptomyces coelicolor partial genome sequence 2.  
XX  
XX antibiotic-producing strain; antibiotic production; scbA gene; ScbR gene;  
KW actinorhodin; undecylprodigiosin; gene; ds.  
XX  
XX Streptomyces coelicolor.  
XX  
XX  
XX Key Location/Qualifiers  
FH complement(1199..2143)  
FT /tag= a  
FT /product= "Streptomyces coelicolor SxBr protein"  
FT /transl_except= (3149..3147, aa:Tyr)  
FT 2261..2908  
FT CDS  
FT /tag= b  
FT /product= "Streptomyces coelicolor SxBr protein"  
FT complement(3024..3812)  
FT /tag= c  
FT /product= "Streptomyces coelicolor SxBr protein"  
FT /transl_except= (3150..3145, aa:Met)  
XX  
XX US2003124644-A1.  
XX  
XX 03-JUL-2003.  
XX  
XX 23-OCT-2001; 2001US-00017471.  
XX  
XX 23-OCT-2000; 2000US-0242561P.  
XX  
XX (TAKA/) TAKANO E.  
PA (BIBB/) BIBB M J.  
XX  
XX Takano E, Bibb MJ;  
XX  
XX WPI, 2003-810983/76.  
DR P-PSDB; ADE14791, ADE14792, ADE14793.  
XX  
XX Modifying an antibiotic-producing strain of Streptomyces coelicolor or  
PT Streptomyces lividans to increase or alter the timing of antibiotic  
PT production in the strain, comprises functionally deleting in the strain  
PT the scbA or ScbR gene.  
XX  
XX Claim 19; Fig 14; 33pp; English.  
XX  
XX
```

CC This invention relates to the novel modification of an antibiotic-  
CC producing strain of Streptomyces coelicolor or Streptomyces lividans to  
CC increase or to alter the timing of antibiotic production in the strain.  
CC The method comprises functionally deleting in the strain the scdA or scbR  
CC gene. The method is useful in increasing and altering the timing of  
CC antibiotic production (especially actinorhodin and undecylprodigiosin) in  
CC Streptomyces species, particularly Streptomyces coelicolor or  
CC Streptomyces lividans. The present sequence is that of a region of the  
CC Streptomyces coelicolor genome, which encodes the scdA, scbB and scbR  
CC proteins and which is related to the invention.  
XX  
SQ Sequence 4346 BP; 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;  
Query Match 80.0%; Score 19.2; DB 10; Length 4346;  
Best Local Similarity 75.0%; Pred. No. 3.7e+02;  
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCCTGSGCCGCGTACCGGAC 24  
Db 1250 GTCCTGATGCCCGGTGACCGCCAC 1273  
RESULT 5  
ABD11456  
ID ABD11456 standard; DNA; 669 BP.  
XX  
AC ABD11456;  
XX  
XX 29-JUL-2004 (first entry)  
XX  
DB Pseudomonas aeruginosa polynucleotide #10060.  
XX  
KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
XX antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
XX US6551795-B1.  
XX  
XX 22-APR-2003.  
XX  
XX 18-FEB-1999; 99US-00252991.  
XX  
XX 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI: 2003-615309/58.  
XX P-PSDB; ABO77885.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 10060; 455bp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences ABD01397-  
XX ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
XX The sequence data for this patent did not form part of the printed

CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 669 BP; 80 A; 230 C; 229 G; 130 T; 0 U; 0 Other;  
Query Match 78.3%; Score 18.8; DB 11; Length 669;  
Best Local Similarity 70.8%; Pred. No. 5.4e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCCTGSGCCGCGTACCGGAC 24  
Db 252 GTCCTGCTTGCCGCGCACCGCCAC 275  
RESULT 6  
ABD11169/c  
ID ABD11169 standard; DNA; 810 BP.  
XX  
XX ABD11169;  
XX  
XX 29-JUL-2004 (first entry)  
XX  
XX  
DE Pseudomonas aeruginosa polynucleotide #9773.  
XX  
KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
XX antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
XX US6551795-B1.  
XX  
XX 22-APR-2003.  
XX  
XX 18-FEB-1999; 99US-00252991.  
XX  
XX 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI: 2003-615309/58.  
XX P-PSDB; ABO77598.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 9773; 455bp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences ABD01397-  
XX ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
XX The sequence data for this patent did not form part of the printed  
XX seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 810 BP; 155 A; 278 C; 275 G; 102 T; 0 U; 0 Other;  
Query Match 78.3%; Score 18.8; DB 11; Length 810;  
Best Local Similarity 70.8%; Pred. No. 5.4e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCTGTGCGCCGTSACSCGSAC 24  
|||||:||||:||||:||||:  
DB 634 GTCTGTGCGCCGCGCACCGCAC 611

RESULT 7  
ADS64127/c  
ADS64127 standard; cDNA; 1100 BP.  
XX  
AC ADS64127;  
XX  
DT 02-DEC-2004 (first entry)  
XX  
DE Bacterial polynucleotide #16114.  
XX  
KW Recombinant DNA construct; transformed plant; improved plant property;  
cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
pathogen tolerance; pest tolerance; plant disease resistance;  
cell cycle pathway modification; plant growth regulator;  
homologous recombination; seed oil yield; protein yield; carbohydrate;  
nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
bacterial polynucleotide; gene; ss.  
XX  
OS Bacteria.  
XX  
PN US2003233675-A1.  
XX  
PD 18-DEC-2003.  
XX  
PF 20-FEB-2003; 2003US-00369493.  
XX  
PR 21-FEB-2002; 2002US-0360039P.  
XX  
PA (CAOY/) CAO Y.  
PA (HINK/) HINKLE G J.  
PA (SLAT/) SLATER S C.  
PA (CHEN/) CHEN X.  
PA (GOLD/) GOLDMAN B S.  
XX  
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
XX  
DR WPI; 2004-061375/06.  
XX  
PT New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.  
XX  
PS Claim 1; SEQ ID NO 39801; 122pp; English.

The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polynucleotide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic

CC format from USPTO at seqdata.uspto.gov/sequence.html.  
XX  
SQ Sequence 1100 BP; 218 A; 393 C; 319 G; 170 T; 0 U; 0 Other;  
XX  
Query Match 78.3%; Score 18.8; DB 13; Length 1100;  
Best Local Similarity 70.8%; Pred. No. 5.4e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCTGTGCGCCGTSACSCGSAC 24  
|||||:||||:||||:||||:  
DB 603 GTCTGTGCGCGGTCACACGAC 580

RESULT 8  
ADS63375/c  
ADS63375 standard; cDNA; 1136 BP.  
XX  
AC ADS63375;  
XX  
DT 02-DEC-2004 (first entry)  
XX  
DE Bacterial polynucleotide #15362.  
XX  
KW Recombinant DNA construct; transformed plant; improved plant property;  
cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
pathogen tolerance; pest tolerance; plant disease resistance;  
cell cycle pathway modification; plant growth regulator;  
homologous recombination; seed oil yield; protein yield; carbohydrate;  
nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
bacterial polynucleotide; gene; ss.  
XX  
OS Bacteria.  
XX  
PN US2003233675-A1.  
XX  
PD 18-DEC-2003.  
XX  
PF 20-FEB-2003; 2003US-00369493.  
XX  
PR 21-FEB-2002; 2002US-0360039P.  
XX  
PA (CAOY/) CAO Y.  
PA (HINK/) HINKLE G J.  
PA (SLAT/) SLATER S C.  
PA (CHEN/) CHEN X.  
PA (GOLD/) GOLDMAN B S.  
XX  
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
XX  
DR WPI; 2004-061375/06.  
XX  
PT New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.  
XX  
PS Claim 1; SEQ ID NO 39049; 122pp; English.

The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or

CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polynucleotide used in  
CC the scope of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification but was obtained in electronic  
CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
SQ Sequence 1136 BP; 223 A; 407 C; 328 G; 178 T; 0 U; 0 Other;  
XX  
Query Match 78.3%; Score 18.8; DB 13; Length 1136;  
Best Local Similarity 70.8%; Pred. No. 5.4e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
XX  
Gy 1 GTCCGTGCGCCGTCGACGAC 24  
Db 639 GTCCGTGCGTCGCGTGACGACGAC 616  
XX  
RESULT 9  
ID ADS63742/c standard; cDNA; 1136 BP.  
XX  
AC ADS63742;  
XX  
DT 02-DEC-2004 (first entry)  
XX  
DE Bacterial polynucleotide #15729.  
XX  
KM Recombinant DNA construct; transformed plant; improved plant property;  
KM cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;  
KM pathogen tolerance; pest tolerance; plant disease resistance;  
KM cell cycle pathway modification; plant growth regulator;  
KM homologous recombination; seed oil yield; protein yield; carbohydrate;  
KM nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KM bacterial polynucleotide; gene; ss.  
XX  
OS Bacteria.  
XX  
PN US2003233675-A1.  
XX  
PD 18-DEC-2003.  
XX  
PF 20-FEB-2003; 2003US-00369493.  
XX  
PR 21-FEB-2002; 2002US-0360039P.  
XX  
XX  
XX (CAOY/) CAO Y.  
XX (HINK/) HINKLE G J.  
XX (SLAT/) SLATER S C.  
XX (CHEN/) CHEN X.  
XX (GOLD/) GOLDMAN B S.  
XX  
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
XX  
DR WPI; 2004-061375/06.  
XX  
PT New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.  
XX  
XX  
PS Claim 1; SEQ ID NO 39416; 122pp; English.  
XX  
XX The invention relates to a recombinant DNA construct comprising a  
XX promoter functional in a plant cell, where the promoter is positioned to  
XX provide for expression of a polynucleotide encoding a polypeptide from a  
XX microbial source. The invention also relates to a transformed plant  
XX comprising the recombinant DNA construct and a method of producing a  
XX transformed plant having an improved property. The plant is a crop plant  
XX such as maize or soybean. The method of producing a transformed plant  
XX having an improved property comprises transforming a plant with the  
XX recombinant DNA construct and growing the transformed plant, where the

CC polynucleotide or polypeptide is useful for improving plant properties.  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polynucleotide used in  
CC the scope of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification but was obtained in electronic  
CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
SQ Sequence 1136 BP; 223 A; 407 C; 328 G; 178 T; 0 U; 0 Other;  
XX  
Query Match 78.3%; Score 18.8; DB 13; Length 1136;  
Best Local Similarity 70.8%; Pred. No. 5.4e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
XX  
Gy 1 GTCCGTGCGCCGTCGACGAC 24  
Db 639 GTCCGTGCGTCGCGTGACGACGAC 616  
XX  
RESULT 10  
ID ABD09558/c standard; DNA; 1338 BP.  
XX  
AC ABD09558;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polynucleotide #8162.  
XX  
KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KM antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
DR WPI; 2003-615309/58.  
XX  
DR P-PSDB; ABO75987.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX  
PS Disclosure; SEQ ID NO 8162; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target

CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)  
CC  
XX  
SQ Sequence 1338 BP; 333 A; 404 C; 347 G; 254 T; 0 U; 0 Other;  
  
Query Match 78.3%; Score 18.8; DB 11; Length 1338;  
Best Local Similarity 70.8%; Pred. No. 5.4e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 GTCTGCTGGCCGCTGACCGCAGC 24  
|||||:|||||:|||||:|||||:|||||:  
Db 999 GTCCGCGCTGCGCGCTGACCCCGAC 976  
  
RESULT 11  
ABD11263/C  
ID ABD11263 standard; DNA; 4248 BP.  
XX  
XX ABD11263;  
AC  
XX  
XX 29-JUL-2004 (first entry)  
DT  
XX  
XX *Pseudomonas aeruginosa* polynucleotide #9867.  
DE  
XX  
XX *Bacterial* infection; gene; ds; *Pseudomonas aeruginosa* infection;  
KM antibacterial.  
XX  
XX *Pseudomonas aeruginosa*.  
OS  
XX  
XX US6551795-B1.  
PN  
XX 22-APR-2003.  
PD  
XX  
XX 18-FEB-1999; 99US-00252991.  
PF  
XX  
XX 18-FEB-1998; 98US-0074788P.  
PR  
XX 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PA  
XX  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
PI  
XX WPI, 2003-615309/58.  
DR  
XX P-PSDB; ABO77692.  
XX  
XX Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,  
XX useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 9867; 455bp; English.  
XX  
XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a *P. aeruginosa* nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-*P. aeruginosa* drugs, as templates for recombinant  
XX production of *P. aeruginosa*-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of *P. aeruginosa*-caused  
XX infection, and in detection of *P. aeruginosa* sequences or other sequences  
XX of *Pseudomonas* species using biochip technology. Sequences ABD01397-  
XX ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
XX The sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)  
XX

SQ Sequence 4248 BP; 832 A; 1408 C; 1389 G; 619 T; 0 U; 0 Other;  
  
Query Match 78.3%; Score 18.8; DB 11; Length 4248;  
Best Local Similarity 70.8%; Pred. No. 5.3e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 GTCTGCTGGCCGCTGACCGCAGC 24  
|||||:|||||:|||||:|||||:|||||:  
Db 975 GTCTGCTGGCCGCTGACCCCGAC 952  
  
RESULT 12  
ACA42624/C  
ID ACA42624 standard; DNA; 4254 BP.  
XX  
XX ACA42624;  
AC  
XX  
XX 19-JUN-2003 (first entry)  
DT  
XX  
XX Prokaryotic essential gene #24281.  
DE  
XX  
XX Antisense; ds; prokaryotic essential gene, cell proliferation;  
XX drug design; gene.  
XX  
XX *Pseudomonas aeruginosa*.  
OS  
XX  
XX WO200277183-A2.  
FN  
XX  
XX 03-OCT-2002.  
PD  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
PF  
XX  
XX 21-MAR-2001; 2001US-00815242.  
PR  
XX 06-SEP-2001; 2001US-00948993.  
PR  
XX 25-OCT-2001; 2001US-0342923P.  
PR  
XX 08-FEB-2002; 2002US-00072851.  
PR  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
PA  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
PI  
XX WPI, 2003-029926/02.  
DR  
XX P-PSDB; ABO38754.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 30494; 1766bp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the



```
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4254 BP; 837 A; 1404 C; 1389 G; 624 T; 0 U; 0 Other;
XX
Query Match      78.3%; Score 18.8; DB 8; Length 4254;
Best Local Similarity 70.8%; Pred. No. 5.3e+02;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
OY 1 GTCCGSGTGGCCSCTMSACSCGAC 24
Db          927 GTCTGTGTCGCCGCACC CGCAC 904

RESULT 13
AAT80415
ID AAT80415 standard; DNA; 13987 BP.
XX
AC AAT80415;
XX
DT 02-MAR-1996 (first entry)
XX
DE Hybrid brmG/cylG ORF1.
XX
KW Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KM platenoidae synthase gene cluster; platenoidae production; smg gene;
KM polyketide; tyllactone synthesis; antibiotic; tylostin; hybrid gene; ss.
OS Streptomyces ambofaciens.
OS Streptomyces fradiae.
XX
FH Key Location/Qualifiers
FT CDS 350..13987
FT /*tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT /note= "ORF1 encodes hybrid protein shown in AAW22611"
PN EP791655-A2.
XX
PD 27-AUG-1997.
XX
PF 19-FEB-1997; 97EP-00301056.
XX
PR 22-FEB-1996; 96US-0012078P.
PA (ELIL ) LILLY & CO ELI.
PI Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
PX MPI: 1997-418046/39.
PY P-PSDB; AAW22611.
PT DNA encoding Streptomyces fradiae tyllactone synthase domain - for
PT production of tylostin-related polyketide compounds.
XX
PS Claim 22; Page 178-197; 220pp; English.
XX
This sequence represents a hybrid gene of the invention. This sequence
CC was created by replacing a EcoRI-ApaI fragment of smg ORF1 with a EcoR-
CC SfiI fragment from tylG ORF1. The position of the nucleotides from each
CC of the two genes is not given in the specification. The smg gene (see
CC AAT80414) was isolated from Streptomyces ambofaciens, and encodes the
CC multi-functional proteins which direct the synthesis of the polyketide
CC platenoidae. Platenoidae is the basic building block of the macrocyclic
CC antibiotic spiramycin. The tylG gene (see AAT80413) is the tyllactone
```

Query Match	78.3%	Score 18.0	DB 2	Length 13987
Best Local Similarity	70.8%	Pred. No. 5.2e+02		
Matches 17	Conservative 5	Mismatches 2	Indels 0	Gaps 0
1	GTCTGTGCGCGGCGTACGCGSAC	24		
2	:     :     :			
3	:     :     :			
4	:     :     :			
5	:     :     :			
6	:     :     :			
7	:     :     :			
8	:     :     :			
9	:     :     :			
10	:     :     :			
11	:     :     :			
12	:     :     :			
13	:     :     :			
14	:     :     :			
15	:     :     :			
16	:     :     :			
17	:     :     :			
18	:     :     :			
19	:     :     :			
20	:     :     :			
21	:     :     :			
22	:     :     :			
23	:     :     :			
24	:     :     :			
25	:     :     :			
26	:     :     :			
27	:     :     :			
28	:     :     :			
29	:     :     :			
30	:     :     :			
31	:     :     :			
32	:     :     :			
33	:     :     :			
34	:     :     :			
35	:     :     :			
36	:     :     :			
37	:     :     :			
38	:     :     :			
39	:     :     :			
40	:     :     :			
41	:     :     :			
42	:     :     :			
43	:     :     :			
44	:     :     :			
45	:     :     :			
46	:     :     :			
47	:     :     :			
48	:     :     :			
49	:     :     :			
50	:     :     :			
51	:     :     :			
52	:     :     :			
53	:     :     :			
54	:     :     :			
55	:     :     :			
56	:     :     :			
57	:     :     :			
58	:     :     :			
59	:     :     :			
60	:     :     :			
61	:     :     :			
62	:     :     :			
63	:     :     :			
64	:     :     :			
65	:     :     :			
66	:     :     :			
67	:     :     :			
68	:     :     :			
69	:     :     :			
70	:     :     :			
71	:     :     :			
72	:     :     :			
73	:     :     :			
74	:     :     :			
75	:     :     :			
76	:     :     :			
77	:     :     :			
78	:     :     :			
79	:     :     :			
80	:     :     :			
81	:     :     :			
82	:     :     :			
83	:     :     :			
84	:     :     :			
85	:     :     :			
86	:     :     :			
87	:     :     :			
88	:     :     :			
89	:     :     :			
90	:     :     :			
91	:     :     :			
92	:     :     :			
93	:     :     :			
94	:     :     :			
95	:     :     :			
96	:     :     :			</

PT production of tylosin-related polyketide compounds.  
 XX  
 XX  
 PS Example 2; Page 110-134; 220pp; English.  
 CC This sequence represents the platenolide synthase gene cluster of the  
 CC invention. This sequence is referred to as the *smg* gene, and was  
 CC isolated from *Streptomyces ambofaciens*. This sequence encodes the multi-  
 CC functional proteins which direct the synthesis of the polyketide  
 CC platenolide. Platenolide is the basic building block of the macrolic  
 CC antibiotic spiramycin. This sequence was used along with the *tylG* gene  
 CC (see AAT80413) to create a hybrid *ORF* sequence (see AAT80415). The *tylG*  
 CC gene is the tylosone synthase gene cluster of the invention. The *tylG*  
 CC sequence was isolated from *Streptomyces fradiae*, and encodes  
 CC multifunctional proteins which direct the synthesis of the polyketide  
 CC tylosone. Tylosone is the basic building block of the antibiotic  
 CC tylosin. The hybrid sequence can be used to transform *S. ambofaciens*  
 CC lacking the *smg* *ORF* sequence, or *S. fradiae* lacking the *tylG* *ORF*  
 CC sequence, so that they can produce polyketides. The DNA sequence can be  
 CC modified so as to alter the type of carboxylic acids incorporated, the  
 CC number of carboxylic acids incorporated and/or the post-condensation  
 CC reactions performed, thereby resulting in novel tylosin-related  
 CC polyketides  
 SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;  
 Query Match 78.3%; Score 18.8; DB 2; Length 44377;  
 Best Local Similarity 70.8%; Pred. No. 5.1e+02;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GTCCGTGTCGCGGTGACGCGCAG 24  
 ID AAT78508  
 ID AAT78508 standard; DNA; 44377 BP.  
 AC AAT78508;  
 XX  
 XX  
 DT 26-FEB-1998 (first entry)  
 XX  
 DE Platenolide synthase gene cluster.  
 XX  
 XX Platenolide synthase gene cluster; platenolide production; *smg* gene;  
 KM multi-functional protein; macrolic antibiotic; spiramycin; ss.  
 XX  
 OS *Streptomyces ambofaciens*.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 350..14002  
 FT /\*tag= a  
 FT /transl\_except= (pos:350..352, aa:Met)  
 FT /note= "ORF1 encodes protein shown in AAW23716"  
 FT 14046..20036  
 FT /\*tag= b  
 FT /note= "ORF2 encodes protein shown in AAW23717"  
 FT 20110..31284  
 FT /\*tag= c  
 FT /transl\_except= (pos:20111..20113, aa:Met)  
 FT /note= "ORF3 encodes protein shown in AAW23718"  
 FT 31329..36071  
 FT /\*tag= d  
 FT /note= "ORF4 encodes protein shown in AAW23719"  
 FT 36155..41830  
 FT /\*tag= e  
 FT /note= "ORF5 encodes protein shown in AAW23720"  
 XX  
 XX  
 PD 27-AUG-1997.  
 XX  
 PF 19-FEB-1997; 97EP-00301066.

XX  
 PR 22-FEB-1996; 96US-0012050P.  
 XX  
 XX  
 PA (ELIL ) LILLY & CO ELI.  
 XX  
 PI Burger SG, Kuhatoos SA, Rao RN, Richardson MA, Rosteck PR;  
 DR WPI; 1997-418047/39.  
 DR P-PSDB; AAW23716, AAW23717, AAW23718, AAW23719, AAW23720.  
 PT DNA encoding *Streptomyces ambofaciens* platenolide synthase domain - for  
 PT production of spiramycin-related polyketide antibiotics.  
 XX  
 PS Claim 9; Page 8-33; 81pp; English.  
 XX  
 XX This sequence represents the platenolide synthase gene cluster of the  
 CC invention. This sequence is referred to as the *smg* gene, and was  
 CC isolated from *Streptomyces ambofaciens*. This sequence encodes the multi-  
 CC functional proteins which direct the synthesis of the polyketide  
 CC platenolide. Platenolide is the basic building block of the macrolic  
 CC antibiotic spiramycin. The DNA can be used to produce compounds  
 CC exhibiting antibiotic activity based on the platenolide structure,  
 CC including specifically the macrolic antibiotic spiramycin and spiramycin  
 CC analogues and derivatives. Modifications of the platenolide synthase DNA  
 CC sequence can be made so as to change the number and type of carboxylic  
 CC acids incorporated into the growing polyketide chain and to change the  
 CC kind of post-condensation processing that is conducted  
 SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;  
 Query Match 78.3%; Score 18.8; DB 2; Length 44377;  
 Best Local Similarity 70.8%; Pred. No. 5.1e+02;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GTCCGTGTCGCGGTGACGCGCAG 24  
 ID AAT78508  
 ID AAT78508 standard; DNA; 44377 BP.  
 AC AAT78508;  
 XX  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Rice gene, SEQ ID 4622.  
 XX  
 XX Plant; bacterial infection; fungal infection; viral infection; rice;  
 KW gene; ds.  
 XX  
 OS *Oryza sativa*.  
 XX  
 PN WO2003000898-A1.  
 XX  
 PD 03-JAN-2003.  
 XX  
 PF 22-JUN-2001; 2001WO-IB001105.  
 XX  
 PR 22-JUN-2001; 2001WO-IB001105.  
 XX  
 XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 PA Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 PI Katagiri F, Quan S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;  
 XX WPI; 2003-175290/17.  
 DR  
 XX  
 XX Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.

```

XX  Claim 6; SEQ ID NO 4622; 899bp; English.
XX
CC  The present invention relates to a method (M1) for identifying genes
CC  involved in plant resistance or response to pathogenic infection. M1
CC  comprises identifying a gene whose expression is significantly altered in
CC  the incompatible interaction of plant gene expression relative to
CC  expression of the gene in an uninfected plant, in a mutant plant that
CC  does not express a gene associated with response to pathogenic infection,
CC  or in a corresponding incompatible or compatible interaction. (M1) is
CC  useful for conferring resistance to resistance or tolerance to a plant to
CC  bacterial, fungal or viral infection. The present sequence was used to
CC  illustrate the invention.
XX
SQ  Sequence 3168 BP; 509 A; 1015 C; 1090 G; 553 T; 0 U; 1 Other;
    76.7%; Score 18.4; DB 8; Length 3168;
    Best Local Similarity 72.7%; Pred. No. 7.6e+02;
    Matches 16; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY  3 CCTGTGGCCGCGTACCGCAGAC 24
    |||||:||||:||||:||||:
Db  3061 CCTGTGGCCGCGTACCGCAGAC 3040

RESULT 17
AAT93095
ID  AAT93095 standard; cDNA; 24379 BP.
XX
AC  AAT93095;
XX
DT  11-MAY-1998 (first entry)
XX
DE  Streptomyces frenolicin gene cluster.
XX
KW  Frenolicin; antibiotic; feed additive; anticoccidial; coccidiostatic;
KW  efflux pump; butyrate starter synthase; polyketide synthase; PKS;
KW  hemiketalase; ketoreductase; cyclase; dehydrase; ketoreductase;
KW  hydroxylase; Streptomyces roseofulvus; ds.
XX
OS  Streptomyces sp.
XX
FH  Location/Qualifiers
FT  CDS
    /tag= a
    /product= "80 kDa non-membrane protein"
    /note= "gene A (specifically claimed)"
    2945..3916
    /tag= b
    /product= "membrane protein"
    /note= "gene B (specifically claimed)"
    4020..4844
    /tag= c
    /product= "protein with 6 membrane-spanning domains"
    /note= "gene C (specifically claimed)"
    4841..6415
    /tag= d
    /product= "ATP-binding component of ABC transporter"
    /note= "gene D (specifically claimed)"
    6533..7183
    /tag= e
    /product= "unknown non-membrane protein"
    /note= "gene E (specifically claimed)"
    7344..8897
    /tag= f
    /product= "putative efflux pump"
    /note= "gene F (specifically claimed)"
    9164..10012
    /tag= g
    /product= "transcription activator"
    /note= "gene G (specifically claimed)"
    complement(10105..10621)
    /tag= h
FT  CDS
    /product= "translationally coupled to gene I"
    /note= "gene H (specifically claimed)"
    complement(10618..11628)
    /tag= i
    /product= "homologue of fabH"
    /note= "gene I (specifically claimed)"
    11809..12066
    /tag= j
    /product= "acyl carrier protein"
    /note= "gene J (specifically claimed)"
    complement(12154..13209)
    /tag= k
    /product= "acyltransferase"
    /note= "gene K (specifically claimed)"
    13409..14686
    /tag= l
    /product= "PKS ketosynthase subunit"
    /note= "gene L (specifically claimed)"
    14767..16047
    /tag= m
    /product= "PKS chain length factor"
    /note= "gene M (specifically claimed)"
    16120..16371
    /tag= n
    /product= "acyl carrier protein used by the PKS"
    /note= "gene N (specifically claimed)"
    complement(16453..16935)
    /tag= o
    /product= "putative hemiketal dehydrase"
    /note= "gene O (specifically claimed)"
    17088..17903
    /tag= p
    /product= "ketoreductase related to actIII"
    /note= "gene P (specifically claimed)"
    17903..18898
    /tag= q
    /product= "cyclase/dehydrase related to act VII"
    /note= "gene Q (specifically claimed)"
    18895..19839
    /tag= r
    /product= "cyclase/dehydrase related to actIV"
    /note= "gene R (specifically claimed)"
    complement(19990..20907)
    /tag= s
    /product= "oxidoreductase"
    /note= "gene S (specifically claimed)"
    complement(20904..22094)
    /tag= t
    /product= "quinone-forming hydroxylase"
    /note= "gene T (specifically claimed)"
    complement(22505..22179)
    /tag= u
    /product= "not specified"
    /note= "gene U"
FT  CDS
    EP806480-A2.
    12-NOV-1997.
    97EP-00107329.
    02-MAY-1997;
    96US-0016753P.
    07-MAY-1996;
    96US-0016753P.
    (HOFF) HOFFMANN LA ROCHE & CO AG F.
    Reeves CD, Soliday CL;
    WI; 1997-538619/50.
    P-PSDB: AAM34199, AAM34200, AAM34201, AAM34202, AAM34203, AAM34204,
    AAM34205, AAM34206, AAM34207, AAM34208, AAM34209, AAM34210, AAM34211,
    AAM34212, AAM34213, AAM34214, AAM34215, AAM34216, AAM34217, AAM34218,
    AAM34219.

```



CC containing the frenolicin gene cluster DNA sequence operably connected to  
CC an expression control sequence; (2) a host cell transformed by the above  
CC vector; (3) a protein coded by the above DNA sequence; (4) a method for  
CC the preparation of frenolicin or a biosynthetic intermediate for it in  
CC which the above cell is cultured and frenolicin or its biosynthetic  
CC intermediate is isolated from the culture or the cell; (5) a method for  
CC the preparation of frenolicin B by oxidizing frenolicin, and (6) a method  
CC for the preparation of a feed composition by mixing frenolicin with other  
CC components. Frenolicin B is useful as an antibiotic  
XX

SO Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;

Query Match 75.0%; Score 18; DB 2; Length 24379;  
Best Local Similarity 75.0%; Pred. No. 1.1e+03;  
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 5 TGCTGCGCCGCTGACCGGAC 24  
|||:|||||:|||||:|||||:  
DB 7005 TGCTGCGCGGTGACCGGCAC 7024

## RESULT 19

ABD02704/C  
ID ABD02704 standard; DNA; 375 BP.

AC ABD02704;

DT 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polynucleotide #1308.

KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
antibacterial.

OS Pseudomonas aeruginosa.

PN US6551795-B1.

PD 22-APR-2003.

PF 18-FEB-1999; 99US-00252991.

PR 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;

DR WPI; 2003-615309/58.

DR P-PSDB; ABO69133.

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.

PS Disclosure; SEQ ID NO 1308; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at

CC segdata.uspto.gov/sequence.html  
XX  
SQ Sequence 375 BP; 74 A; 130 C; 138 G; 33 T; 0 U; 0 Other;

Query Match 74.2%; Score 17.8; DB 11; Length 375;  
Best Local Similarity 69.6%; Pred. No. 1.4e+03;  
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCTGCGCCGCTGACCGGGA 23  
|||:|||||:|||||:|||||:  
DB 117 GTCTGCTGACGCTGACCGGA 95

RESULT 20  
ID ABD02553 standard; DNA; 681 BP.

AC ABD02553;

DT 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polynucleotide #1157.

KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
antibacterial.

OS Pseudomonas aeruginosa.

PN US6551795-B1.

PD 22-APR-2003.

PF 18-FEB-1999; 99US-00252991.

PR 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;

DR WPI; 2003-615309/58.

DR P-PSDB; ABO68982.

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.

PS Disclosure; SEQ ID NO 1157; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC segdata.uspto.gov/sequence.html

XX Sequence 681 BP; 88 A; 242 C; 244 G; 107 T; 0 U; 0 Other;

Query Match 74.2%; Score 17.8; DB 11; Length 681;  
Best Local Similarity 69.6%; Pred. No. 1.3e+03;  
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GTCTGTGCGCCGATACCGGA 23  
Db 610 GTCTGTGCGATGACCGGA 632

RESULT 21  
ADP04866  
ID ADP04866 standard; cDNA, 1455 BP.  
XX  
XX ADP04866;  
AC  
XX  
XX 29-JUL-2004 (first entry)  
DT  
XX

DE Sea squirt cDNA with tissue specific expression in development Seq 461.  
XX  
XX gene; ss; sea squirt; regeneration medicine; gene therapy;  
KM cell proliferation; differentiation; reproduction;  
KM environmental measurement; water survey.  
XX  
XX Ciona intestinalis.  
OS

PN JP2004057129-A.  
XX  
XX 26-FEB-2004.  
PD  
XX 31-JUL-2002; 2002JP-00222593.  
PF  
XX 31-JUL-2002; 2002JP-00222593.  
PR  
XX  
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
PA  
XX

DR WPI; 2004-287079/27.  
DR P-PSDB; ADP04867.  
XX  
XX Novel gene cluster which is specifically expressed in tissue or organ  
PT during development phase of sea squirt, useful for elucidation of  
PT mechanism of development of tissue or organ of sea squirt.  
XX  
XX Claim 2; SEQ ID NO 461; 1846bp; Japanese.

CC This invention relates to novel genes and the encoded proteins thereof  
CC that are derived from the sea squirt Ciona intestinalis. Specifically, it  
CC refers to those genes that are expressed in the tissues or organs of the  
CC sea squirt during its developmental phase. The present invention  
CC describes the identification of these genes as useful for elucidation of  
CC the mechanism of development and hence for developing regeneration  
CC medicines and gene therapy techniques. Accordingly, they can be used in  
CC the research of various genetic diseases, as well as the analysis of cell  
CC proliferation, differentiation and reproduction. Furthermore, such  
CC compositions can be useful for environmental measurements and water  
CC surveys, particularly for sea water surveys, and also for the preparation  
CC of transformed sea squirt for improving edibility of sea squirt such as  
CC Halocynthia roretzi. This polynucleotide sequence is a sea squirt cDNA  
CC sequence that exhibits tissue specific expression during development,  
CC given in an exemplification of the invention.  
XX  
XX SQ Sequence 1455 BP; 412 A; 258 C; 364 G; 421 T; 0 U; 0 Other;

Query Match 74.2%; Score 17.8; DB 12; Length 1455;  
Best Local Similarity 69.6%; Pred. No. 1.3e+03;  
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
Oy 1 GTCTGTGCGCCGATACCGGA 23  
Db 984 GTTCTGTGCGCGATACCGGA 1006

RESULT 22  
ABD07378  
ID ABD07378 standard; DNA; 492 BP.  
XX  
XX ABD07378;  
XX

DT 29-JUL-2004 (first entry)  
XX  
XX Pseudomonas aeruginosa polynucleotide #5982.  
DE  
XX  
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KM antibacterial.  
XX  
XX Pseudomonas aeruginosa.  
OS

PN US6551795-B1.  
XX  
XX 22-APR-2003.  
PD  
XX 18-FEB-1999; 99US-00252991.  
PF

XX 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PA  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
PI WPI; 2003-615309/58.  
DR P-PSDB; ABO73807.  
XX

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 5982; 455bp; English.

CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC segata.uspto.gov/sequence.html  
XX  
XX SQ Sequence 492 BP; 57 A; 174 C; 162 G; 99 T; 0 U; 0 Other;

Query Match 72.5%; Score 17.4; DB 11; Length 492;  
Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
Oy 2 TCCTGTGCGCCGATACCGGC 22  
Db 232 TCCTGTGCGCGATACCGGC 252

RESULT 23  
ABD01735/C  
ID ABD01735 standard; DNA; 654 BP.  
XX  
XX ABD01735;  
AC  
XX  
XX 29-JUL-2004 (first entry)  
DT

DE Pseudomonas aeruginosa polynucleotide #339.  
XX  
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KM antibacterial.  
XX  
XX Pseudomonas aeruginosa.  
OS

XX US6551795-B1.  
PN 22-APR-2003.  
XX  
XX 18-FEB-1999; 99US-00252991.  
PF  
XX 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PA  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX WPI; 2003-615309/58.  
XX P-PSDB; ABO68164.  
DR  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 339; 455bp; English.  
PS  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using bioclip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC the sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 654 BP; 89 A; 214 C; 259 G; 92 T; 0 U; 0 Other;  
Query Match 72.5%; Score 17.4; DB 11; Length 654;  
Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TCCTGCTGGCCGCTGACGCG 22  
DB 548 TCCTGCTGGCCGCTGACGCG 528  
RESULT 24  
AAS54125  
ID AAS54125 standard; DNA; 744 BP.  
XX  
XX AAS54125;  
AC  
XX 13-FEB-2002 (first entry)  
DT  
XX Pseudomonas aeruginosa DNA for cellular proliferation protein #256.  
DE  
XX Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;  
KW antibacterial; drug design.  
XX  
XX Pseudomonas aeruginosa.  
OS  
XX MO200170955-A2.  
PN  
XX 27-SEP-2001.  
PD  
XX 21-MAR-2001; 2001WO-US009180.  
PF  
XX 21-MAR-2000; 2000US-0191078P.  
PR

PR 23-MAY-2000; 2000US-0206848P.  
PR 26-MAY-2000; 2000US-0207727P.  
PR 23-OCT-2000; 2000US-0242578P.  
PR 27-NOV-2000; 2000US-0253625P.  
PR 22-DEC-2000; 2000US-0257931P.  
PR 16-FEB-2001; 2001US-0269308P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
PA  
PI Haselbeck R, Ohlsen KL, Zykkind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX WPI; 2001-611495/70.  
XX P-PSDB; AAU36266.  
DR  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids.  
PT  
XX  
XX Claim 27; SEQ ID NO 7762; 511bp; English.  
PS  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the genes,  
CC their use in the discovery of novel antibiotics, the essential genes  
CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence encodes an essential prokaryotic  
CC cellular proliferation protein. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 744 BP; 103 A; 281 C; 257 G; 103 T; 0 U; 0 Other;  
Query Match 72.5%; Score 17.4; DB 4; Length 744;  
Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TCCTGCTGGCCGCTGACGCG 22  
DB 602 TCCTGCTGGCCGCTGACGCG 622  
RESULT 25  
ACA42298  
ID ACA42298 standard; DNA; 744 BP.  
XX  
XX ACA42298;  
AC  
XX 19-JUN-2003 (first entry)  
DT  
XX Prokaryotic essential gene #23955.  
DE  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
XX Pseudomonas aeruginosa.  
OS  
XX MO200277183-A2.  
PN  
XX 03-OCT-2002.  
PD  
XX 21-MAR-2002; 2002WO-US009107.  
PF  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR

25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX MPI; 2003-029926/02.  
DR P-PSDB; ABU38428.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 30168; 1766bp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 744 BP; 103 A; 281 C; 257 G; 103 T; 0 U; 0 Other;  
XX  
Query Match 72.5%; Score 17.4; DB 8; Length 744;  
Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
OY 2 TCCTGCTGCGCCGCTGACGCGS 22  
DB 602 TCCTGCTGCGCCGCTGACGCGS 622  
XX  
RESULT 26  
ABD01726  
ID ABD01726 standard; DNA; 759 BP.  
XX  
AC ABD01726;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polynucleotide #330.  
XX  
KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KM antibacterial.

XX  
OS Pseudomonas aeruginosa.  
XX  
XX US6551795-B1.  
XX  
XX 22-APR-2003.  
XX  
XX 18-FEB-1999; 99US-00252991.  
XX  
XX 18-FEB-1998; 98US-0074788P.  
XX  
XX 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX MPI; 2003-615309/58.  
XX  
XX P-PSDB; ABO68155.  
XX  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 330; 455bp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences ABD01397-  
XX ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
XX The sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX segdata.uspto.gov/sequence.html  
XX  
SQ Sequence 759 BP; 108 A; 283 C; 262 G; 106 T; 0 U; 0 Other;  
XX  
Query Match 72.5%; Score 17.4; DB 11; Length 759;  
Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
OY 2 TCCTGCTGCGCCGCTGACGCGS 22  
DB 617 TCCTGCTGCGCCGCTGACGCGS 637  
XX  
RESULT 27  
ACH97630  
ID ACH97630 standard; DNA; 843 BP.  
XX  
AC ACH97630;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Klebsiella pneumoniae polynucleotide segid 3425.  
XX  
XX Recombinant expression vector; transcription regulatory element;  
XX Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.  
XX  
XX Klebsiella pneumoniae.  
XX  
XX US6610836-B1.  
XX  
XX 26-AUG-2003.  
XX  
XX 27-JAN-2000; 2000US-00489039.



XX 29-JAN-1999; 99US-0117747P.  
PR (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX  
PI Breton GL, Osborne M;  
XX  
DR WPI; 2003-895346/82.  
DR P-PSDB; ABO64079.  
XX  
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.  
XX  
XX Disclosure; SEQ ID NO 3425; 932pp; English.  
XX  
CC The invention describes a new isolated nucleic acid encoding a Klebsiella  
CC pneumoniae polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid, operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella  
CC pneumoniae polypeptide of the invention  
XX  
SQ Sequence 843 BP; 166 A; 274 C; 253 G; 150 T; 0 U; 0 Other;  
XX  
Query Match 72.5%; Score 17.4; DB 11; Length 843;  
Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
OY 4 CTGCTGCGCCGTCGACCGGAC 24  
DB 337 CTGCTGCGCCGTCGACCGGAC 357  
XX  
RESULT 28  
ABD01715  
ID ABD01715 standard; DNA; 1026 BP.  
XX  
AC ABD01715;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polynucleotide #319.  
XX  
KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KM antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Rubenfeld MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI; 2003-615309/58.  
DR P-PSDB; ABO68144.  
XX  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 319; 455pp; English.  
XX  
CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and

CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 1026 BP; 148 A; 400 C; 330 G; 147 T; 0 U; 1 Other;  
XX  
Query Match 72.5%; Score 17.4; DB 11; Length 1026;  
Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
OY 2 TCCTGTGCGCCGTCGACCGG 22  
DB 78 TCCTGTGCGCCGTCGACCGG 98  
XX  
RESULT 29  
ABG90070  
ID ABG90070 standard; DNA; 1212 BP.  
XX  
AC ABG90070;  
XX  
DT 01-OCT-2002 (first entry)  
XX  
DE M. capsulatus gene #55 for DNA array.  
XX  
KM Micro array; gene; ds; differential expression; gene expression.  
XX  
OS Methylococcus capsulatus.  
XX  
PN WO200255655-A2.  
XX  
PD 18-JUL-2002.  
XX  
PF 14-JAN-2002; 2002WO-N0000019.  
XX  
PR 12-JAN-2001; 2001NO-00000235.  
PR 12-JAN-2001; 2001NO-00000239.  
XX  
PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.  
PA (TIGR-) TIGR.  
XX  
XX Birkeland NK, Bidhammer I, Jonassen I, Jensen HB, Lien T;  
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;  
PI Salzberg SL;  
XX  
XX WPI; 2002-557818/59.  
DR  
XX  
XX Novel DNA array useful for determining differential expression of  
XX Methylococcus capsulatus genes, comprises polynucleotides or  
XX oligonucleotides representative for a selective number of Methylococcus  
XX capsulatus genes.  
XX  
XX Claim 19; Page 76-77; 678pp; English.  
XX  
XX The invention relates to a novel DNA array giving a representation of a  
XX number of Methylococcus capsulatus genes. The method of the invention is  
XX useful for determination of the differential expression of the genes of  
XX M. capsulatus, and for studying gene expression on a genomic scale and in  
XX gene expression assays of M. capsulatus genes. The sequences shown in  
XX ABG90016-ABG91855 represent M. capsulatus genes for use in arrays of the  
XX invention

SO Sequence 1212 BP; 233 A; 375 C; 384 G; 219 T; 0 U; 1 Other;

Query Match 72.5%; Score 17.4; DB 6; Length 1212;  
 Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
 Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 4 CTGCTGCGCCGTCGTCGCGAC 24  
 |||:|||||:|:|:|  
 DB 823 CTGCTGCGCGTCGTCGCGAC 843

RESULT 30

ABD07361  
 ID ABD07361 standard; DNA; 1578 BP.

AC ABD07361;

DT 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polynucleotide #5965.

XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
 KW antibacterial.

XX Pseudomonas aeruginosa.

XX US6551795-B1.

XX 22-APR-2003.

XX 18-FEB-1999; 99US-00252991.

XX 18-FEB-1998; 98US-0074788P.

XX 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX WPI: 2003-615309/58.

XX P-PsDB; ABO73790.

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.

PS Disclosure; SEQ ID NO 5965; 455bp; English.

CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences  
 CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html

XX Sequence 1578 BP; 192 A; 537 C; 530 G; 319 T; 0 U; 0 Other;

Query Match 72.5%; Score 17.4; DB 11; Length 1578;  
 Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
 Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 TCCTGCTGCGCCGTCGTCGCGS 22  
 |||:|||||:|:|:|

DB 434 TCCTGCTGCGCGATCACCGC 454

Search completed: July 20, 2005, 16:11:57  
 Job time : 420.857 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 14:42:10 ; Search time 1136.57 Seconds  
(without alignments)  
1023.187 Million cell updates/sec

Title: US-10-017-471B-2

Perfect score: 24  
Sequence: 1 gtcctgctgcgcgtacacgcgac 24

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1\_0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

GenEmbl:.\*  
1: gb\_ba:.\*  
2: gb\_hcg:.\*  
3: gb\_in:.\*  
4: gb\_om:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_pl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_str:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	22	91.7	1215	1 AB001608 Streptomy
C 2	21	87.5	6252	1 SVU24659 Streptomyc
C 3	20.4	85.0	2234	1 AB001683 Streptomyc
C 4	20	83.3	93902	8 AP003767 Oryza sat
C 5	20	83.3	234050	1 AL627282 Salmoneil
C 6	20	83.3	300620	1 AE016848 Salmoneil
C 7	19.6	81.7	1203	1 STMARSAA
C 8	19.6	81.7	12070	1 AB011413 Streptomyc
C 9	19.2	80.0	3118	8 AK110337 Oryza sat
C 10	19.2	80.0	6727	1 SC0007731 Streptomyc
C 11	19.2	80.0	131711	2 AP003954 Oryza sat
C 12	19.2	80.0	145772	8 AE004269 Oryza sat
C 13	19.2	80.0	230850	1 SC0393127 Streptomyc
C 14	18.8	78.3	1851	10 BC032284 Mus muscu
C 15	18.8	78.3	1886	10 AB036747 Mus muscu
C 16	18.8	78.3	1901	10 AB036748 Mus muscu
C 17	18.8	78.3	1904	10 AB036746 Mus muscu
C 18	18.8	78.3	1919	10 AB036749 Mus muscu
C 19	18.8	78.3	8992	1 SV117268 Streptomyc

C 20	18.8	78.3	11004	1 AE012334 Xanthomon
C 21	18.8	78.3	13513	1 AF506520 Streptomyc
C 22	18.8	78.3	15356	1 AE004867 Pseudomon
C 23	18.8	78.3	34398	1 AY258138 Pseudomon
C 24	18.8	78.3	50550	7 AF271693 Mycobacte
C 25	18.8	78.3	94058	2 AC141024 Rattus no
C 26	18.8	78.3	110000	2 AP006618 Continuation (14 o
C 27	18.8	78.3	146585	8 CNS080CAM_13
C 28	18.8	78.3	152186	2 AC084284
C 29	18.8	78.3	162995	10 AC084302
C 30	18.8	78.3	193050	2 AL646062 Ralstonia
C 31	18.8	78.3	199271	2 AC142243 Mus muscu
C 32	18.8	78.3	229737	2 AC097884 Rattus no
C 33	18.8	78.3	303855	1 AE017230 Mycobacte
C 34	18.4	76.7	1850	1 TSU17342
C 35	18.4	76.7	3168	6 AX654752
C 36	18.4	76.7	5367	1 SCABAGC2
C 37	18.4	76.7	9090	1 AY392413 Streptomyc
C 38	18.4	76.7	9316	1 AF077869 Streptomyc
C 39	18.4	76.7	11069	1 AE008208 Agrobacte
C 40	18.4	76.7	11293	1 AE009406 Agrobacte
C 41	18.4	76.7	12358	1 AE012437 Xanthomon
C 42	18.4	76.7	24720	1 AJ632203 Streptomyc
C 43	18.4	76.7	110000	1 AP006618_18
C 44	18.4	76.7	110000	1 AP006618_59
C 45	18.4	76.7	139773	2 AC067806
C 46	18.4	76.7	140226	2 AC010092
C 47	18.4	76.7	343473	9 BX640451 Bordetell
C 48	18.2	75.8	84248	1 AP001783 Homo sapi
C 49	18.2	75.8	140140	2 AC151647 Dasyatis n
C 50	18.2	75.8	153699	2 AC144285 Macaca mu
C 51	18.2	75.8	165005	2 AC009940 Homo sapi
C 52	18.2	75.8	250792	2 AC096277 Rattus no
C 53	18	75.0	25306	1 AF058302 Streptomyc
C 54	18	75.0	110000	2 AP006490_2
C 55	17.8	74.2	309	12 AF657179 Continuation (3 of
C 56	17.8	74.2	1455	3 AK116370 Synthetic
C 57	17.8	74.2	1521	1 AMU03061 AK116370 Ciona int
C 58	17.8	74.2	3573	1 AB016763 Streptomyc
C 59	17.8	74.2	11057	1 AE012461 Xanthomon
C 60	17.8	74.2	12633	1 AE013939 Yersinia
C 61	17.8	74.2	16268	1 AE004753 Pseudomon
C 62	17.8	74.2	21562	1 PPS79894 Pseudomon
C 63	17.8	74.2	51975	9 AC108475 Homo sapi
C 64	17.8	74.2	98734	9 AC013277 Homo sapi
C 65	17.8	74.2	109528	1 AF040570 Amycolato
C 66	17.8	74.2	110000	1 AE017282_09 Continuation (10 o
C 67	17.8	74.2	110000	1 AE017282_13 Continuation (14 o
C 68	17.8	74.2	110000	1 AP006618_25 Continuation (26 o
C 69	17.8	74.2	110000	1 BX571965_14 Continuation (15 o
C 70	17.8	74.2	110000	1 BX936398_38 Continuation (39 o
C 71	17.8	74.2	110000	8 CR382129_07 Continuation (8 of
C 72	17.8	74.2	205544	2 CR847834 CR847834 Danio rer
C 73	17.8	74.2	208050	1 AJ414145 Yersinia
C 74	17.8	74.2	216614	1 AB088824 Streptomyc
C 75	17.8	74.2	226349	2 AC140865 AC140865 Homo sapi
C 76	17.8	74.2	238543	2 CR456635 CR456635 Danio rer
C 77	17.8	74.2	280558	1 AE017301 Thermus t
C 78	17.8	74.2	289555	1 AP005947 Bradyrhiz
C 79	17.8	74.2	290333	1 AE017140 Yersinia
C 80	17.8	74.2	299800	1 AP005028 Streptomyc
C 81	17.8	74.2	310174	1 AE016870 Pseudomon
C 82	17.6	73.3	452	1 AF010105 Unidentif
C 83	17.6	73.3	460	1 AY651310 Unculture
C 84	17.6	73.3	498	1 AY214755 Unculture
C 85	17.6	73.3	1419	1 AF134587 Streptomyc
C 86	17.6	73.3	1482	1 AB117717 Unculture
C 87	17.6	73.3	9971	1 AE000820 Methanoba
C 88	17.6	73.3	45603	1 AY196994 Streptomyc
C 89	17.6	73.3	105841	8 AC092697 Oryza sat
C 90	17.6	73.3	110000	1 BX571966_08 Continuation (9 of
C 91	17.6	73.3	110000	1 CP000011_08 Continuation (9 of
C 92	17.6	73.3	162215	9 AC004896 Homo sapi

C 93	17.6	73.3	194361	8	AC092389	AC092389 Oryza sat	C 166	17.2	71.7	4511	6	CO729197	CO729197 Sequence
C 94	17.6	73.3	201201	2	AC150433	AC150433 Colobus g	C 167	17.2	71.7	4511	9	AB023202	AB023202 Homo sapi
C 95	17.6	73.3	300029	8	AB017115	AB017115 Oryza sat	C 168	17.2	71.7	5137	9	AK090454	AK090454 Homo sapi
C 96	17.6	73.3	300100	1	SCO939123	SCO939123 Streptomy	C 169	17.2	71.7	5344	9	HSN807425	HSN807425 Homo sapi
C 97	17.6	73.3	302070	1	AP005823	AP005823 Corynebact	C 170	17.2	71.7	5373	3	SADNMTW5	SADNMTW5 Streptomy
C 98	17.6	73.3	310550	1	SCO939113	SCO939113 Streptomy	C 171	17.2	71.7	6523	1	AB017045	AB017045 Paracoccu
C 99	17.4	72.5	843	6	AR386696	AR386696 Sequence	C 172	17.2	71.7	6653	1	AY162971	AY162971 Micromono
C 100	17.4	72.5	1212	6	AR363420	AR363420 Sequence	C 173	17.2	71.7	7055	1	ASP457162	ASP457162 Arthrobac
C 101	17.4	72.5	1688	1	STMSUBCB	STMSUBCB S.griiseolus	C 174	17.2	71.7	7403	8	AF047459	AF047459 Chlamydo
C 102	17.4	72.5	1998	6	AR363414	AR363414 Sequence	C 175	17.2	71.7	7860	1	SKZ86111	SKZ86111 Streptomy
C 103	17.4	72.5	2351	9	CO842135	CO842135 Sequence	C 176	17.2	71.7	7980	1	MXU20669	MXU20669 Myxococcu
C 104	17.4	72.5	2351	9	AK125166	AK125166 Homo sapi	C 177	17.2	71.7	8568	8	AF449619	AF449619 Phycophth
C 105	17.4	72.5	9828	1	AB004468	AB004468 Pseudomon	C 178	17.2	71.7	10136	1	AF546156	AF546156 Micromono
C 106	17.4	72.5	10202	1	AB004483	AB004483 Pseudomon	C 179	17.2	71.7	10932	1	BCPYR	BCPYR B.caldolyti
C 107	17.4	72.5	11842	1	AB005053	AB005053 Halobacte	C 180	17.2	71.7	11092	1	AB005703	AB005703 Caulobact
C 108	17.4	72.5	11886	1	AB004638	AB004638 Pseudomon	C 181	17.2	71.7	12675	1	AB001878	AB001878 Deinococc
C 109	17.4	72.5	47344	9	AF061479	AF061479 Homo sapi	C 182	17.2	71.7	12908	1	AE004934	AE004934 Pseudomon
C 110	17.4	72.5	97556	9	AC079781	AC079781 Homo sapi	C 183	17.2	71.7	12995	1	AB011857	AB011857 Xanthomon
C 111	17.4	72.5	99549	2	AC092862_3	Continuation (4 of	C 184	17.2	71.7	13056	1	AE004713	AE004713 Pseudomon
C 112	17.4	72.5	110000	2	AE017282_02	Continuation (3 of	C 185	17.2	71.7	13268	6	AX587577	AX587577 Sequence
C 113	17.4	72.5	110000	2	AC119727_1	Continuation (2 of	C 186	17.2	71.7	13528	1	AE014340	AE014340 Brucella
C 114	17.4	72.5	110121	9	AL590989	AL590989 Human DNA	C 187	17.2	71.7	13556	1	AE009600	AE009600 Brucella
C 115	17.4	72.5	117296	9	AC092865	AC092865 Homo sapi	C 188	17.2	71.7	13857	6	AR338584	AR338584 Sequence
C 116	17.4	72.5	135431	9	AC106847	AC106847 Homo sapi	C 189	17.2	71.7	14219	1	AB070957	AB070957 Streptomy
C 117	17.4	72.5	143717	2	AC083933	AC083933 Homo sapi	C 190	17.2	71.7	14467	9	CVU84760	CVU84760 Chromatium
C 118	17.4	72.5	166246	2	AF271408	AF271408 Homo sapi	C 191	17.2	71.7	17019	9	EX936374	EX936374 Human DNA
C 119	17.4	72.5	167576	2	AC147297	AC147297 Pan trogl	C 192	17.2	71.7	18977	1	SAR7932	AB070932 Streptomy
C 120	17.4	72.5	167875	2	AC083906	AC083906 Homo sapi	C 193	17.2	71.7	27521	1	AB112586	AB112586 Streptomy
C 121	17.4	72.5	167914	2	AC055870	AC055870 Homo sapi	C 194	17.2	71.7	30000	6	AX250263	AX250263 Sequence
C 122	17.4	72.5	167675	2	AC016978	AC016978 Homo sapi	C 195	17.2	71.7	34869	1	AF324838	AF324838 Streptomy
C 123	17.4	72.5	185636	2	AC145875	AC145875 Pan trogl	C 196	17.2	71.7	36602	6	CO878999	CO878999 Sequence
C 124	17.4	72.5	281450	1	AB005032	AB005032 Streptomy	C 197	17.2	71.7	36748	7	AY539836	AY539836 Burkholde
C 125	17.4	72.5	303206	2	AC122945	AC122945 Rattus no	C 198	17.2	71.7	38846	2	AC145717	AC145717 Homo sapi
C 126	17.4	72.5	332889	2	AC098450	AC098450 Rattus no	C 199	17.2	71.7	39568	2	AC014823	AC014823 Drosophil
C 127	17.4	72.5	339650	1	SCO939108	SCO939108 Streptomy	C 200	17.2	71.7	39753	2	AC149318	AC149318 Phakopsor
C 128	17.2	71.7	318	9	HSN17868	Y17868 Homo sapien	C 201	17.2	71.7	46777	9	EX247883	EX247883 Human DNA
C 129	17.2	71.7	453	12	AY657216	AY657216 Synthetic	C 202	17.2	71.7	51915	9	AP005273	AP005273 Homo sapi
C 130	17.2	71.7	498	9	AF424846	AF424846 Macaca mu	C 203	17.2	71.7	57339	9	AL928593	AL928593 Human DNA
C 131	17.2	71.7	603	6	AX196096	AX196096 Sequence	C 204	17.2	71.7	82746	1	AF453501	AF453501 Actinobyn
C 132	17.2	71.7	708	8	AU066535	AU066535 Chlamydom	C 205	17.2	71.7	88421	6	AX417445	AX417445 Sequence
C 133	17.2	71.7	825	6	CO594282	CO594282 Sequence	C 206	17.2	71.7	96153	2	CR755972	CR755972 Homo sapi
C 134	17.2	71.7	861	3	AY118988	AY118988 Drosophil	C 207	17.2	71.7	96335	9	EX679671	EX679671 Human DNA
C 135	17.2	71.7	1415	1	THR295159	THR295159 Thermus t	C 208	17.2	71.7	100267	9	HSNHC36A	HSNHC36A Homo sapien
C 136	17.2	71.7	1500	5	BC065586	BC065586 Dantio rer	C 209	17.2	71.7	102750	9	AC004465	AC004465 Homo sapi
C 137	17.2	71.7	1632	6	AX654521	AX654521 Sequence	C 210	17.2	71.7	103450	1	AF440781	AF440781 Streptomy
C 138	17.2	71.7	1826	1	ABN1FH	ABN1FH X51500 Acospirillu	C 211	17.2	71.7	104677	9	CR753803	CR753803 Human DNA
C 139	17.2	71.7	1934	9	BC009843	BC009843 Homo sapi	C 212	17.2	71.7	106117	2	AP000628	AP000628 Homo sapi
C 140	17.2	71.7	1968	9	BC002605	BC002605 Homo sapi	C 213	17.2	71.7	107413	4	AL773560	AL773560 Pig DNA
C 141	17.2	71.7	2047	8	AK110881	AK110881 Oryza sat	C 214	17.2	71.7	109519	6	AX195929	AX195929 Sequence
C 142	17.2	71.7	2070	6	CO879003	CO879003 Sequence	C 215	17.2	71.7	109528	1	AF040570	AF040570 Amycolato
C 143	17.2	71.7	2127	6	CO586451	CO586451 Sequence	C 216	17.2	71.7	109646	9	HSNHC78S22	HSNHC78S22 Streptomy
C 144	17.2	71.7	2189	6	AK834331	AK834331 Sequence	C 217	17.2	71.7	110000	1	AE016822_06	Continuation (7 of
C 145	17.2	71.7	2189	9	AK096721	AK096721 Homo sapi	C 218	17.2	71.7	110000	1	AE017282_00	AE017282 Methyloco
C 146	17.2	71.7	2454	8	AP022816	AP022816 Chlamydom	C 219	17.2	71.7	110000	1	AE017282_13	Continuation (14 o
C 147	17.2	71.7	2648	1	SCSPRAGE	X94190 S.coelicolo	C 220	17.2	71.7	110000	1	AE017282_16	Continuation (17 o
C 148	17.2	71.7	2688	6	CO879005	CO879005 Sequence	C 221	17.2	71.7	110000	1	AP006618_17	Continuation (18 o
C 149	17.2	71.7	2712	6	PDEPHAAC	DA3764 Paracoccus	C 222	17.2	71.7	110000	1	AP006618_23	Continuation (24 o
C 150	17.2	71.7	2712	6	AX105326	AX105326 Sequence	C 223	17.2	71.7	110000	1	AP006618_43	Continuation (44 o
C 151	17.2	71.7	2825	6	CO594281	CO594281 Sequence	C 224	17.2	71.7	110000	1	AP006618_44	Continuation (45 o
C 152	17.2	71.7	2867	8	AB015139	AB015139 Chlamydom	C 225	17.2	71.7	110000	1	AP006618_52	Continuation (53 o
C 153	17.2	71.7	2888	6	E34218	E34218 Chlorophyll	C 226	17.2	71.7	110000	1	AP006840_01	Continuation (2 of
C 154	17.2	71.7	2963	9	BC017259	BC017259 Homo sapi	C 227	17.2	71.7	110000	1	AP006840_17	Continuation (18 o
C 155	17.2	71.7	3042	1	AB091059	AB091059 Gluconace	C 228	17.2	71.7	110000	1	AP006840_17	Continuation (23 o
C 156	17.2	71.7	3146	1	AY348315	AY348315 Streptomy	C 229	17.2	71.7	110000	2	AP006840_22	Continuation (4 of
C 157	17.2	71.7	3271	3	SD0580406	SD0580406 Subterfies	C 230	17.2	71.7	110000	2	AP006496_3	Continuation (14 o
C 158	17.2	71.7	3292	1	BSA11805	AB071166 Bacillus	C 231	17.2	71.7	113647	9	AC026425_3	Continuation (3 of
C 159	17.2	71.7	3542	1	AB071166	AB071166 Gluconace	C 232	17.2	71.7	115292	2	AC120117	Continuation (18 o
C 160	17.2	71.7	3835	9	HSNHC71B	X60189 Human MHC c	C 233	17.2	71.7	119151	2	AP000655	Continuation (23 o
C 161	17.2	71.7	4050	1	STWILVBCN	L33268 Streptomyce	C 234	17.2	71.7	120628	9	AC012153	Continuation (14 o
C 162	17.2	71.7	4133	1	AB054887	AB054887 Streptomy	C 235	17.2	71.7	129811	9	HSNHC34F7	Continuation (14 o
C 163	17.2	71.7	4142	9	AB124552	AB124552 Homo sapi	C 236	17.2	71.7	130141	9	AL513320	Continuation (14 o
C 164	17.2	71.7	4201	1	AY588479	AY588479 Pseudomon	C 237	17.2	71.7	135638	9	AF484556	Continuation (14 o
C 165	17.2	71.7	4210	1	MSGILVB	L49392 Mycobacteri	C 238	17.2	71.7	138992	9	AL662828	Continuation (14 o

239	17.2	71.7	142560	2	CR753845
C 240	17.2	71.7	145299	2	AF228728
C 241	17.2	71.7	146249	2	AF359826
C 242	17.2	71.7	146436	2	AC108759
C 243	17.2	71.7	147891	8	AP004693
C 244	17.2	71.7	149303	8	AC087412
C 245	17.2	71.7	150925	9	AC113607
C 246	17.2	71.7	150925	9	AC125282
C 247	17.2	71.7	155503	2	AC084347
C 248	17.2	71.7	155159	9	AL591916
C 249	17.2	71.7	160487	5	BX005126
C 250	17.2	71.7	163785	3	BX547938
C 251	17.2	71.7	163279	3	AC104508
C 252	17.2	71.7	165079	2	AC024517
C 253	17.2	71.7	167300	2	AC021394
C 254	17.2	71.7	169259	9	AP006288
C 255	17.2	71.7	169740	2	AF186193
C 256	17.2	71.7	169856	2	AC144890
C 257	17.2	71.7	170859	10	AC128952
C 258	17.2	71.7	172249	2	CR753839
C 259	17.2	71.7	176267	4	AC148660
C 260	17.2	71.7	176991	4	BX296515
C 261	17.2	71.7	180559	9	AL645922
C 262	17.2	71.7	181210	9	AC011742
C 263	17.2	71.7	181858	9	AC084856
C 264	17.2	71.7	182705	3	AC091208
C 265	17.2	71.7	184418	2	AC130644
C 266	17.2	71.7	186024	2	AC084425
C 267	17.2	71.7	186335	2	BX927323
C 268	17.2	71.7	187081	10	AC124421
C 269	17.2	71.7	188700	2	AC021603
C 270	17.2	71.7	189050	1	AL646066
C 271	17.2	71.7	189050	1	AL646069
C 272	17.2	71.7	189424	9	AC148664
C 273	17.2	71.7	193915	2	AC091457
C 274	17.2	71.7	199204	2	AC131298
C 275	17.2	71.7	203613	10	AL591404
C 276	17.2	71.7	207206	9	AC090260
C 277	17.2	71.7	210018	2	AC015557
C 278	17.2	71.7	213050	1	AL646079
C 279	17.2	71.7	213201	2	BX927095
C 280	17.2	71.7	215210	2	CR749744
C 281	17.2	71.7	218445	2	AC083959
C 282	17.2	71.7	219682	2	AC150692
C 283	17.2	71.7	224640	2	AC125590
C 284	17.2	71.7	250762	2	AL662840
C 285	17.2	71.7	257843	2	AC095980
C 286	17.2	71.7	272233	2	AC106289
C 287	17.2	71.7	273285	1	AB017304
C 288	17.2	71.7	273769	2	AC099070
C 289	17.2	71.7	295150	1	SC0939125
C 290	17.2	71.7	298300	1	AP005025
C 291	17.2	71.7	299050	1	SC0939119
C 292	17.2	71.7	299300	1	AP005026
C 293	17.2	71.7	299986	1	AB017240
C 294	17.2	71.7	300511	1	AB016775
C 295	17.2	71.7	300861	1	AB016777
C 296	17.2	71.7	301399	1	AB017233
C 297	17.2	71.7	302007	1	SC0939132
C 298	17.2	71.7	302325	1	AB017236
C 299	17.2	71.7	302325	1	AB017236
C 300	17.2	71.7	302898	1	AB017238

## ALIGNMENTS

RESULT 1  
LOCUS AB001608/c 1215 bp DNA linear BCT 22-NOV-1997  
DEFINITION Streptomyces virginiae DNA for Barx, complete cds.  
ACCESSION AB001608  
VERSION AB001608.1 GI:2641955

KEYWORDS Barx.  
SOURCE Streptomyces virginiae  
ORGANISM Streptomyces virginiae  
REFERENCE 1 (bases 1 to 1215)  
AUTHORS Kinoshita, H., Ipposhi, H., Okamoto, S., Nakano, H., Nihira, T. and Yamada, Y.  
TITLE Butyrolactone autoregulator receptor protein (Barx) as a transcriptional regulator in Streptomyces virginiae  
JOURNAL J. Bacteriol. 179 (22), 6986-6993 (1997)  
MEDLINE 98037495  
PUBMED 9371444  
REFERENCE 2 (bases 1 to 1215)  
AUTHORS Kinoshita, H.  
TITLE Direct Submission  
JOURNAL Submitted (06-MAR-1997) Hiroshi Kinoshita, Osaka University, Department of Engineering, Yamadaoka 2-1, Suita, Osaka 565, Japan (E-mail:kinoshita@biochem.bio.eng.osaka-u.ac.jp, Tel:+81-6-879-7433, Fax:+81-6-879-7432)  
FEATURES  
source location/Qualifiers  
1. .1215  
/organism="Streptomyces virginiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1961"  
244. .1128  
/gene="barx"  
244. .1128  
/gene="barx"  
/codon\_start=1  
/transl\_table=1  
/product="Barx"  
/protein\_id="BAA23611.1"  
/db\_xref="GI:2641956"  
/translation="MTSTVDRILVHRAVAVELITGMSPTAENRPLATTQWPRASVFTPVNGCDPILASBITRQVGTLLSHAFRCVSFDDQLMNDLHSHVPRBOGVGAAPRDLELDVICSIDIRRRGRRLAGKRYEVTLICGGQVATAGAAFDCTSPVYTORLDRYGA TGVRPLQAPLAPASVGRFLTTDVLSATERPLEMQRVDEQHPVLDHPVDHVGWLT MESARPLAQAIDPSRPFLLPTMRSEFSRYALDRPCWIQEPLPADNGDRQVRVIGH QDDTVFSCILIGRGAAB"

## ORIGIN

Query Match 91.7%; Score 22; DB 1; Length 1215;  
Best Local Similarity 79.2%; Pred. No. 4,4e+02;  
Matches 19; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCTGTGCGCGGTGACGCGSAC 24  
Db 1077 GTCTGTGCGCGGTGACGCGGAC 1054

RESULT 2  
SVU24659/c 6252 bp DNA linear BCT 07-AUG-2003  
LOCUS SVU24659 Streptomyces venezuelae jad gene cluster, complete sequence.  
DEFINITION U24659  
ACCESSION U24659  
VERSION U24659.2 GI:16445340  
KEYWORDS Streptomyces venezuelae  
SOURCE Streptomyces venezuelae  
ORGANISM Streptomyces venezuelae  
REFERENCE 1 (bases 2899 to 6252)  
AUTHORS Yang, K., Han, L. and Vining, L.C.  
TITLE Regulation of jadomycin B production in Streptomyces venezuelae  
JOURNAL J. Bacteriol. 177 (21), 6111-6117 (1995)  
MEDLINE 96042086  
PUBMED 7592375  
REFERENCE 2 (bases 2899 to 6252)  
AUTHORS Yang, K., Han, L., Vining, L.C. and He, J.Y.  
TITLE Participation of jadR in the regulation of jadomycin B production in Streptomyces venezuelae ISP 5230

JOURNAL Unpublished  
 REFERENCE 3 (bases 2899 to 6252)  
 AUTHORS Yang, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-APR-1995) Keqian Yang, Biology, Dalhousie University,  
 Halifax, Nova Scotia B3H 401, Canada  
 REFERENCE 4 (bases 1 to 6252)  
 AUTHORS Wang, L. and Vining, L.C.  
 TITLE Control of growth, secondary metabolism and sporulation in  
 Streptomyces venezuelae ISP5230 by jadW1), a member of the afsA  
 family of gamma-butyrolactone regulatory genes  
 JOURNAL Microbiology (Reading, Engl.) 143 (Pt 8), 1991-2004 (2003)  
 MEDLINE 22787062  
 PUBMED 12904539  
 REFERENCE 5 (bases 1 to 6252)  
 AUTHORS Wang, L. and Vining, L.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-OCT-2001) Biology, Dalhousie University, 1355 Oxford,  
 Halifax, Nova Scotia B3H 401, Canada  
 REMARK Sequence update by submitter  
 COMMENT On Oct 26, 2001 this sequence version replaced gi:886035.  
 FEATURES  
 source location/Qualifiers  
 1..6252  
 /organism="Streptomyces venezuelae"  
 /mol\_type="genomic DNA"  
 /strain="ISP5230"  
 /db\_xref="taxon:54571"  
 /note="wild type"  
 195..1124  
 /gene="jadW1"  
 /note="similar to Barx and AfsA"  
 195..1124  
 /codon\_start=1  
 /transl\_table=11  
 /product="JadW1"  
 /protein\_id="AAB36582.2"  
 /db\_xref="GI:16445341"  
 /translation="MHTSRGEFVRADPADIIPTDWTOLRONRFSVSARVACLSL  
 LSRAGARHPMLVARTIPETSMVLVAHELGVLDOPVWMDLSYSGADEALVLDGLS  
 DVTYVVCSDITRRGSRRLNLTYYVLRDPDLATGSGTARCTALAYRRRGSRME  
 ALGRPEPLIPGVHPVGRFARTEVDVLAENKPDQMLRVNTAHTTLRRRNDHYFCM  
 VILLEAARQAATATGSAAYLPTDLVSFLRYVELDSPCMIEASVPTDPSTTTIKVT  
 GHGDSPVRCITLSPSRRLSVATAGLDTRLAG"  
 1139..2086  
 /gene="jadW2"  
 1139..2086  
 /gene="jadW2"  
 1139..2086  
 /function="Transcriptional regulation"  
 /note="dehydrogenase; repressor"  
 /codon\_start=1  
 /transl\_table=11  
 /product="JadW2"  
 /protein\_id="AAL23835.1"  
 /db\_xref="GI:16445342"  
 /translation="MGPRIILITGAGFGIGRVAATAAREPAHYRLLAGPLTPAA  
 AAGPRTETAVGDLRDPASLRACGEVDVLIHCASAIIGDAELAVNDEGTNLVDA  
 VAAAGRTIYGPRLRTRGPRPRAAPGTLPIAPSGSYSTPAAAGRRVMDGCVTLR  
 PHVVGTDGRLPLGLVRLAALPEPLDGGSSLSHSLIDVSLAGLLGAAISPAIEPG  
 PYYVDHPEPVPGDLIAAESLITRTGLRGAVALDPTARRLLAGLPFAGHLLMLAT  
 DMFPADRPWREIGCDPPGLPRGSRTRTGTDSSTIP"  
 2109..2873  
 /note="dehydrogenase/reductases; regulator for jadomycin B  
 biosynthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /product="JadW3"  
 /protein\_id="AAL23836.1"  
 /db\_xref="GI:16445343"  
 /translation="MSALAGRTAIVGASRGIGRGIAERLATDGAIVAVHYSNKA  
 LRTVELTERRGGAFAFAELGAPDVVDVTFYALDGLERGAAREFDILVNNAAAG  
 SRIHQLTTEVDRLEPAIVNAKPLFLVQGLRLRGGRIVINISSATGAPDSDVTY  
 AATKGAVDMTALAKELGPRGITVNAVAPGYIATDMNARRRATPEASAAIAMSIVN

gene  
 CDS  
 RIGTPADVGRGPRVSDERWITGVYDATTGATL"  
 complement (2953..3543)  
 /gene="jadR2"  
 /complement (2953..3543)  
 /gene="jadR2"  
 /function="transcriptional regulation"  
 /note="repressor"  
 /citation=11  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="JadR2"  
 /protein\_id="AAB36583.1"  
 /db\_xref="GI:886037"  
 /translation="MTQKQERATRTDALISAREFDEHGVALAKLSAISGACVSG  
 ALHFHFNKVAAYEIDASTTLRTARIYVHOSNNAQNLADITHALARLVREDVVR  
 AGFRLSCSOLCGTDNLROMESCVCQGLAEADDEGLASDIGGQDLARTIVATIG  
 LEALCRNGEMLSPGVTGTMLTLPVVAAPGRSP"  
 4192..4896  
 /gene="jadR1"  
 4192..4896  
 /gene="jadR1"  
 /function="transcriptional regulation"  
 /note="response regulator"  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="JadR1"  
 /protein\_id="AAB36584.1"  
 /db\_xref="GI:886038"  
 /translation="MTVDHIGEGGPTTAAGRVLYVESCGKEGGLVQGLRRGHGV  
 DRVETGGALQAYEADLVLIIDELPDLGLEVCGRIGRSGDVPVIAVTSIGSLDV  
 LGLOAGADRYELVKPYGFRLEMAKMEAWRRARPASPAKTIITRGPDLHIDAAQVTL  
 GQVDLTFRKEDLLYLASHPDVTIPIKQLMQGWGDSWRRRTVDTVSSILRNLTGAS  
 DWVITIRGVFRFGG"  
 ORIGIN  
 Query Match 87.5%; Score 21; DB 1; Length 6252;  
 Beet Local Similarity 78.3%; Pred. No. 8.2e+02;  
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCGTGTCGCCGSGTSCGCGSA 23  
 |||||:|||||:|||||:|||||:  
 Db 1034 GTCTCGTGCGCGCGTACCGCGA 1012  
 |||||:|||||:|||||:|||||:  
 RESULT 3  
 LOCUS AB001683 2234 bp DNA linear BCT-22-AUG-1997  
 DEFINITION Streptomyces sp. gene for Farx, Farx, complete cds.  
 ACCESSION AB001683  
 VERSION AB001683.1 GI:2342428  
 KEYWORDS Farx; Farx.  
 SOURCE Streptomyces sp.  
 ORGANISM Streptomyces sp.  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycinae; Streptomycetaceae; Streptomyces.  
 REFERENCE 1 (sites)  
 AUTHORS Waki, M., Nihira, T. and Yamada, Y.  
 TITLE Cloning and characterization of the gene (farx) encoding the  
 receptor for an extracellular regulatory factor (IM-2) from  
 Streptomyces sp. strain FRI-5  
 JOURNAL J. Bacteriol. 179 (16), 5131-5137 (1997)  
 MEDLINE 97405912  
 PUBMED 9260956  
 REFERENCE 2 (bases 1 to 2234)  
 AUTHORS Waki, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-MAR-1997) Miyoko Waki, Osaka University, Graduate  
 School of Engineering, Department of Biotechnology, Yamadaoka 2-1,  
 Suita, Osaka 565, Japan (E-mail:waki.yam@stu.bio.eng.osaka-u.ac.jp,  
 Tel:+81-6-879-7433, Fax:+81-6-879-7432)  
 FEATURES  
 location/Qualifiers

```

source
1. .2234
/organism="Streptomyces sp."
/mol_type="genomic DNA"
/strain="FRI-5"
/db_xref="taxon:1931"
440..1315
/gene="fark"
440..1315
/gene="fark"
/clone_start=1
/transl_except=1
/transl_table=1
/product="Fark"
/protein_id="BAA21858.1"
/db_xref="GI:2342429"
/translation="MVRTSTACVLTDMORLDDARFVTAAPMPLSHAFPTVDGY
DPLMCAETIOIATLGHAEPAVFGHOFVMDLSVSVBELRLVATVDLAT
CVELKRRAGRLSGIYEAVVRDGGVATGASVTCSPAVYGRIREHVLTPHRL
PLTRPAAPQSVARLSPFDVVISPLDREKRLRDVTHNPVLPFDHVDHVFQMTLMEA
RQAASALGRPSFMPVAGVAFKRYVELDAPCVIESRLFDVFGAEEVAVVTGONG
ELTFVGTVTASVSGY"
1500..2165
/gene="fara"
1500..2165
/gene="fara"
/codon_start=1
/transl_except=1
/transl_table=1
/product="Fara"
/protein_id="BAA21859.1"
/db_xref="GI:2342430"
/translation="MAEQVRAIRTRQALSLAAAFVPERGVQATTSIELTVAGVTG
ALYFHSKEDLAQGVLTQWEDLLPERPAKQEVVDVWLHTRTPMVAQVR
LSLDVNAAGLDRSAPFRNVWDKFTDLKKAQAGELLPHVVPARTAVITGAGVGS
MSQALTRHQLDGRVNLKRLHMSIAPSVLASLHGESAEVIEAQLAREQAD
EED."

gene
CDS

ORIGIN
Query Match 85.0%; Score 20.4; DB 1; Length 2234;
Best Local Similarity 75.0%; Pred. No. 1.6e+03;
Matches 18; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCTGTCGCGCGTACGCGGAC 24
1264 GTTCGTGTCGCGCGTACGCGGAC 1241

RESULT 4
AP003767/c 93902 bp DNA linear PLN 23-APR-2004
LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,
DEFINITION PAC clone: P0036F10, complete sequence.
ACCESSION AP003767
VERSION AP003767.2 GI:46518322
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehharctoideae; Oryzaceae; Oryza.
1
REFERENCE
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Oryza sativa nipponbare (GMS) genomic DNA, chromosome 6, PAC
clone: P0036F10
JOURNAL Published Only in Database (2001)
2 (bases 1 to 93902)
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2001) Takuji Sasaki, National Institute of
Agricultural Sciences, Rice Genome Research Program, Kamondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsaseki@nias.affrc.go.jp, URL: http://rtp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)

```

```

COMMENT On Apr 22, 2004 this sequence version replaced gi:14517641.
FEATURES
source
1. .93902
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/Chromosome="6"
/clone="P0036F10"

ORIGIN
Query Match 83.3%; Score 20; DB 8; Length 93902;
Best Local Similarity 77.3%; Pred. No. 1.3e+03;
Matches 17; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCTGTCGCGCGTACGCGGAC 24
27007 CCTGTCGCGCGTACGCGGAC 26986

RESULT 5
AL627282/c 234050 bp DNA linear BCT 04-JUN-2003
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
DEFINITION complete chromosome, segment 18/20.
ACCESSION AL627282 AL513382
VERSION AL627282.1 GI:16505159
KEYWORDS
SOURCE Salmonella enterica subsp. enterica serovar Typhi
ORGANISM Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 234050)
REFERENCE
AUTHORS Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,
Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G.,
Sebatina, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,
Conerton, P., Cronin, A., Davies, P., Davies, R.M., Dowd, L., White, N.,
Farrell, J., Felwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S.,
Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P.,
Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
Stevens, K., Whitehead, S. and Barrall, B.G.
TITLE Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
JOURNAL Nature 413 (6858), 848-852 (2001)
21534947
PUBMED 11677608
REFERENCE
AUTHORS Parkhill, J.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).
Location/Qualifiers
1. .234050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/mol_type="genomic DNA"
/strain="CT18"
/db_xref="taxon:90370"
92..463
/gene="rplN"
/Note="synonym: STY4368"
92..463
/gene="rplN"
/Note="Orthologue of E. coli rplN (RL14_ECOLI); Pasta hit
to RL14_ECOLI (123 aa), 98% identity in 123 aa overlap"
/codon_start=1

```

```
misc_feature
/translate=508 ribosomal subunit protein L14"
/product_id="CAD09156.1"
/db_xref="GI:16505160"
/db_xref="GOA:O8XG78"
/db_xref="UniProt/TREMBL:O8XG78"
/translation="MIOGOTMLNVADNSGARVMICIKYLGGSHRRYAGVGDIIKITIK
EAIIPGKVKKGQVLAQVAVVRYTKGVRRPDGSVIRFDGNACVILNNSEQPIGTRIIFGP
VTRLEINKEFMKILSLAPEVL"
92. .460
/gene="rplN"
/note="Pfam match to entry PF00238 Ribosomal L14,
Ribosomal protein L14p/L23e, score 277.30, E-value 2e-79"
269. .349
/gene="rplN"
/note="PS00049 Ribosomal protein L14 signature"
474. .788
/gene="rplX"
/note="synonym: STY4369"
474. .788
/gene="rplX"
/note="Orthologue of E. coli rplX (RL24 ECOLI); Fastra hit
to RL24 ECOLI (103 aa), 99% identity in 103 aa overlap"
/codon_start=1
/translate=508 ribosomal subunit protein L24"
/product_id="CAD09157.1"
/db_xref="GI:16505161"
/db_xref="GOA:O8Z1X8"
/db_xref="UniProt/TREMBL:O8Z1X8"
/translation="MAAKIRRDDEVIVLTGDKGRGKGVNVLSSGKVIIVGINLVRK
HOKPVPALNPGGIVYEKAALQVSNVALFNTATGKADRVGRFEDKKVRRPFKSNSET
IK"
486. .656
/gene="rplX"
/note="Pfam match to entry PF00467 Ribosomal_L24, KOM
motif, score 89.90, E-value 5.1e-23"
495. .548
/gene="rplX"
/note="PS01108 Ribosomal protein L24 signature"
803. .1342
/gene="rplE"
/note="synonym: STY4370"
803. .1342
/gene="rplE"
/note="Orthologue of E. coli rplE (RL5 ECOLI); Fastra hit
to RL5 ECOLI (178 aa), 99% identity in 178 aa overlap"
/codon_start=1
/translate=508 ribosomal subunit protein L5"
/product_id="CAD09158.1"
/db_xref="GI:16505162"
/db_xref="GOA:P37436"
/db_xref="UniProt/Swiss-Prot:P37436"
/translation="MAKLHDYKDEVNKLMTFENYNSVMQVPRVEKITLNMVGEAL
ADKLLDNAAADLTALISQKPLITARRASVAFKIRQSGPICKVTLGERWEEFER
LITIAVPRIIRDFRGLSAXSFDSGRGNVSMGVBRQIIFPEIDYDKVDRVRLGITTTTTA
KSDDEGRALLAADFPPFRK"
872. .1042
/gene="rplE"
/note="Pfam match to entry PF00281 Ribosomal_L5, Ribosomal
protein L5, score 120.00, E-value 4.5e-32"
971. .1021
/gene="rplE"
/note="PS00358 Ribosomal protein L5 signature"
1052. .1336
/gene="rplE"
/note="Pfam match to entry PF00673 Ribosomal_L5_C,
ribosomal L5p family C-terminus, score 205.50, E-value
8.2e-58"
1280. .1333
/gene="rplE"
/note="PS01156 TonB-dependent receptor proteins signature
```

```
gene
2"
1357. .1662
/gene="rpsH"
/note="synonym: STY4371"
1357. .1662
/gene="rpsH"
/note="Orthologue of E. coli rpsH (RS14 ECOLI); Fastra hit
to RS14 ECOLI (100 aa), 98% identity in 100 aa overlap"
/codon_start=1
/translate=308 ribosomal subunit protein S14"
/product_id="CAD09159.1"
/db_xref="GI:16505163"
/db_xref="GOA:O8XEM0"
/db_xref="UniProt/Swiss-Prot:O8XEM0"
/translation="MAQSMKAREVKRVALADKYFARAEIKAIISDVNAITDEDRNA
VLKIQTLPRDSSPQRQNRRCQGTGRPHAFILKFGILSRIKYREAAIRGEIDGLKQASW"
1357. .1659
/gene="rpsH"
/note="Pfam match to entry PF00253 Ribosomal_S14,
Ribosomal protein S14p/S29e, score 177.60, E-value
1.3e-53"
1543. .1611
/gene="rpsH"
/note="PS00527 Ribosomal protein S14 signature"
1696. .2088
/gene="rpsH"
/note="synonym: STY4372"
1696. .2088
/gene="rpsH"
/note="Orthologue of E. coli rpsH (RS8 ECOLI); Fastra hit
to RS8 ECOLI (129 aa), 100% identity in 129 aa overlap"
/codon_start=1
/translate=308 ribosomal subunit protein S8"
/product_id="CAD09160.1"
/db_xref="GI:16505164"
/db_xref="GOA:P02361"
/db_xref="UniProt/Swiss-Prot:P02361"
/translation="MSMQDPIADMLTRIRNGQAKAVTMPSSKLTQVAIANVLKEEG
FIEDFKVGDTPKELETLTKYFGKAVVESIQVSRPQLRIYKKDELPMKMAIGLGA
VSTSKGVMTRARQAQGLGEIICVYA"
1708. .2085
/gene="rpsH"
/note="Pfam match to entry PF00410 Ribosomal_S8, Ribosomal
protein S8, score 248.20, E-value 3.6e-75"
1993. .2046
/gene="rpsH"
/note="PS00053 Ribosomal protein S8 signature"
2101. .2634
/gene="rplF"
/note="synonym: STY4373"
2101. .2634
/gene="rplF"
/note="Orthologue of E. coli rplF (RL6 ECOLI); Fastra hit
to RL6 ECOLI (176 aa), 99% identity in 176 aa overlap"
/codon_start=1
/translate=508 ribosomal subunit protein L6"
/product_id="CAD09161.1"
/db_xref="GI:16505165"
/db_xref="GOA:O8XFL6"
/db_xref="UniProt/Swiss-Prot:O8XFL6"
/db_xref="GOA:O8XFL6"
/translation="MSRVAKAPVVPAVDVQKINGQVITIKKNGELTRTLNDAVEVK
HADNALFGPRDGVADQWAGQGTARALLNSVIVTSGFTKKQLQGVGYRAAYKGVY
VNLSTLGSHPVDHQLPAGITACEPTQREIYVKGADKQVIGQVADLAAVRRPEEYKKG
GVRYADEVRYTKAEKKK"
2131. .2631
/gene="rplF"
/note="Pfam match to entry PF00347 Ribosomal_L6, Ribosomal
protein L6, score 299.10, E-value 5.4e-86"
2560. .2586
/gene="rplF"
```





```

/translation="MKRIAFVSTPAPHSASGREGIDALLATSLATXGVFPISDGV
FOLLPGQPDVALADYATATFKLPDLYIDOCWICAPSLRERGLGNVFNVAITPLEP
VALRRELGNVILRF"
/gene="yhel"
/3860..4147
/locus_tag="t4056"
3860..4147
/gene="yhel"
/locus_tag="t4056"
/locus_tag="t4056"
/note="corresponds to STY4350 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO71523.1"
/db_xref="GI:29139959"
/translation="MLHTLPHCAGVDPPALLRLKEGDALLLLDQGTVAIEGNRFL
ESLRAPITVYALKKDIDARGIGQISDSVYRVDTFRVRLTVKANQMAM"
4273..4647
/gene="tpsl"
/locus_tag="t4057"
4273..4647
/gene="tpsl"
/locus_tag="t4057"
/note="corresponds to STY4350 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="30S ribosomal subunit protein S12"
/protein_id="AAO71524.1"
/db_xref="GI:29139960"
/translation="MATVQIVRKPRARKAVKSNVPALEACPQKRGVCTRYTTTPRK
PNSALRKQCRVRLTGEVTSYIGEGHNLDEHSVILIRGGRVKDLPGVRHYTVAGAL
DQSGVDRKROASKYGVKRPKA"
4743..5213
/gene="tpsg"
/locus_tag="t4058"
4743..5213
/gene="tpsg"
/locus_tag="t4058"
/note="corresponds to STY4351 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="30S ribosomal subunit protein S7"
/protein_id="AAO71525.1"
/db_xref="GI:29139961"
/translation="MRRRVIGQKILDPKFGSILAKFVNILWVQKKSIVASIV
SALETLAQSGSSELEAFVLENVPRVVEKSRVGSSTYQVPEVPRKALAMR
WIVEARRKGDGMALRLANELSDAADNKGTVAKREDVHMEANKAFAHYRW"
5310..7424
/gene="fusa"
/locus_tag="t4059"
5310..7424
/gene="fusa"
/locus_tag="t4059"
/note="corresponds to STY4352 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="elongation factor G"
/protein_id="AAO71526.1"
/db_xref="GI:29139962"
/translation="MARTPIARVYNIIGTSAHIDAGKTTTTERILPYGVNKHIGEVH
DGAATMDMEQERGITTTSAATTAPFSGAKQYEPPIRINIIDTPGHADPTIEVER
KVLDGAVVVCYAVGQVQPSSTWQANKYVPRIAFNKDRMGANLKVQGIKT
RUGANVPVQLQALGAEBGTVADVVKMAIMWADQCVTEYEDIDPADMDLANEW
HONLIEASAESELMETKLGSELTREBIKOALRQRYLANEIIIVTCGSAFKNGVQ
AMLDAVIDLPSVPVPAINGILDGKDTPARHSDDEPSGALPKIATDFVGNLT
PRRVYSGVNSGDYVANSYKTAERERGRIVQGHANKREBIKEVRAGDIAAALGLDVT

```

Query Match 83.3%; Score 20; DB 1; Length 300620;  
Best Local Similarity 77.3%; Pred. No. 1.1e+03;

```

Matches 17; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Qy 2 TCCTGTGCGCCGTSACSCGA 23
|||||:|||||:|||||:
Db 247494 TCCTGTGCGCCGTCACCGCA 247473

RESULT 7
STWAFSAA 1203 bp DNA linear BCT 26-APR-1993
LOCUS S_grieseus afsa gene encoding a possible A-factor biosynthesis
DEFINITION protein.
M24250
ACCESSION M24250.1 GI:153148
VERSION A-factor biosynthesis.
KEYWORDS Streptomyces grieseus
SOURCE Streptomyces grieseus
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycetaceae; Streptomyces.
1 (bases 1 to 1203)
Horinouchi, S., Suzuki, H., Nishiyama, M. and Beppu, T.
Nucleotide sequence and transcriptional analysis of the
Streptomyces grieseus gene (afsa) responsible for A-factor
biosynthesis
JOURNAL J. Bacteriol. 171 (2), 1206-1210 (1989)
MEDLINE 89123125
PUBMED 2492509
COMMENT Original
FEATURES
source location/Qualifiers
1..1203
/organism="Streptomyces grieseus"
/mol_type="genomic DNA"
/db_xref="taxon:1911"
204..1109
/note="afsa protein"
/codon_start=1
/transl_table=11
/protein_id="AA26693.1"
/db_xref="GI:153149"
/translation="MDAEAEVHPVGIEMVHRTPEDEAPRNWRLGRDPAVEAVLP
HDHPFPAVPGDLDHPILVABAMQAMLAFAHAGYGLGVHFLITTEIDYVCHENLG
VGGPRTIGLSEVFCGDLKMRAGLPAGQGVGWVARGRLATGVAATRFSTPKAYRRA
RGDVPVEISISPTAPVPDAPSPAGARVEDVLTSGTGRGWELELVDRHPITLPORPND
HVPMLLIEARQAACLVAGPAGIVPEARTFRIRYSFSGPCWIGAVVQGADEDTY
TVRVTHGDGEIVTSVLSGPRAG"

CDS
Query Match 81.7%; Score 19.6; DB 1; Length 1203;
Best Local Similarity 79.2%; Pred. No. 3.5e+03;
Matches 19; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GTCCGTGCGCCGTSACSCGAC 24
|||||:|||||:|||||:
Db 1058 GTCCGTGCGCCTGTACCCGAC 1035

RESULT 8
AB011413 12070 bp DNA linear BCT 07-AUG-1998
LOCUS Streptomyces grieseus genes for Orf2, Orf3, Orf4, Orf5, AfsA, Orf8,
DEFINITION partial and complete cds.
AB011413
ACCESSION AB011413.1 GI:3401946
VERSION Orf8; AfsA; Orf5; Orf4; Orf3; Orf2.
KEYWORDS Streptomyces grieseus
SOURCE Streptomyces grieseus
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycetaceae; Streptomyces.
1 (sites)
Umeyama, T.
REFERENCE Open reading frame encoded around afsA gene
JOURNAL Unpublished
TITLE Open reading frame encoded around afsA gene
REFERENCE 2 (bases 1 to 12070)

```



Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusnogi, T., Li, C., Lu, M., Maeda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Naniki, T., Narikawa, R., Nishikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otsu, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Tagami, M., Tagami, T., Tagawa, T., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

**COMMENT**  
This clone is one of the 28K full-length cDNA clones from japonica rice.

URL: [http://cdna01.dna.affrc.go.jp/cDNA/NIAS\\_Rice\\_Full-Length\\_cDNA\\_Project\\_Team\\_Kikuchi,S.,Sato,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,J.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hotta,I.,Kojima,K.,Naniki,T.,Ohneda,E.,Yahagi,W.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.andYamamoto,M.](http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-Length_cDNA_Project_Team_Kikuchi,S.,Sato,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,J.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hotta,I.,Kojima,K.,Naniki,T.,Ohneda,E.,Yahagi,W.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.andYamamoto,M.)

**FEATURES**  
source

1. .3118  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="002-164-F01"

**ORIGIN**

Query Match 80.0%; Score 19.2; DB 8; Length 3118;  
Best Local Similarity 75.0%; Pred. No. 4.3e+03;  
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

1 GTCTGTTGCGGCGTACGCGSAC 24  
|||||:|||||:|||||:|||||  
1442 GTCTGCGAGCGCGTACACGAC 1465

**RESULT 10**  
SC0007731 6727 bp DNA linear BCT 24-AUG-1998  
LOCUS Streptomyces coelicolor scbr gene, scbA gene, ORFs A,B,X & Z.  
ACCESSION AJ007731  
VERSION AJ007731.1 GI:3425857  
KEYWORDS gamma-butyrolactone binding protein; scbA gene; scbr gene.

**SOURCE**  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

Streptomyces coelicolor A3 (2)  
Streptomyces coelicolor A3 (2)  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycinae; Streptomycetaceae; Streptomyces.

Takano, E., Chakraborty, R., Nihira, T., Yamada, Y. and Bibb, M.  
Characterisation of scbr, and scbA genes involved in  
gamma-butyrolactone binding and synthesis in Streptomyces  
coelicolor  
Unpublished  
2 (bases 1 to 6727)

**REFERENCE**  
AUTHORS  
TITLE  
JOURNAL

1. .6727  
/organism="Streptomyces coelicolor A3 (2)"  
/mol\_type="genomic DNA"  
/strain="M145"  
/db\_xref="taxon:100226"  
37. .969  
/gene="orfA"  
37. .969  
/gene="orfB"  
/codon\_start=1  
/transl\_table=1  
/product="histidine kinase"  
/protein\_id="CA07625.1"  
/db\_xref="GI:3425858"  
/db\_xref="GOA:O86849"  
/db\_xref="UniProt/TREMBL:O86849"  
/translation="MGTPSPLENRGGLIPCSIDRPSRSRTVERHFCRTPARHD  
LAGHRRGRFETLQDGLSRFRGARTGTGTAHPCGADGRDPAAGVARDRGDRD  
HVMERRRLAEVDEDELGTALSTAHNHLIAETGTPHIDARSHLEMAVARRL  
TGELOATVPEPLAOAVGEFAATRPPTGTEVRIPTTGEBRLISPCREELFAVARRL  
HNATHHMAADVYTTLTFRTRKMAAGIYDQGVGRDAVAVPGHARGLRSMTRIED  
VGMRLIVSGPAGTHIDVHLPLPRKVSIAPR1"  
complement(1000. .1854)  
/gene="orfA"  
complement(1000. .1854)  
/gene="orfB"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="CA07626.1"  
/db\_xref="GI:3425859"  
/db\_xref="UniProt/TREMBL:O86850"  
/translation="WILVTGVYGVGRVVAHLARTGPRVTLARPERVYTGENVY  
CAGEYADPGLDRALNGVRAFLVTSNTEPDDERFAAAREAGVRHLVKLSLAVGE  
PGADDFITRRQRENERRAVRESGLDWTNPARTFMSNTLSMPAIRSDGVRLAGTSP  
VCAVDPRIIAVAVALTRPGEGRAVYLSGPEAITRAOCTARLSEVLSASQODTRRLRV  
DTSHTPLRPFNDHVPGLLLEAARQACLVTPGAPFVPSIGTRFVAYAFDSPCWI  
OATVRPGPAAGLTIVRYTGHQDGLVPLTITLSGAFSG"  
3032. .3679  
/gene="scbr"

**FEATURES**  
source

1. .3118  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="002-164-F01"

**ORIGIN**

Query Match 80.0%; Score 19.2; DB 8; Length 3118;  
Best Local Similarity 75.0%; Pred. No. 4.3e+03;  
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

1 GTCTGTTGCGGCGTACGCGSAC 24  
|||||:|||||:|||||:|||||  
1442 GTCTGCGAGCGCGTACACGAC 1465

**RESULT 10**  
SC0007731 6727 bp DNA linear BCT 24-AUG-1998  
LOCUS Streptomyces coelicolor scbr gene, scbA gene, ORFs A,B,X & Z.  
ACCESSION AJ007731  
VERSION AJ007731.1 GI:3425857  
KEYWORDS gamma-butyrolactone binding protein; scbA gene; scbr gene.

```

CDS
3032..3679
/gene="ecbR"
/codon_start=1
/evidence=experimental
/translation="gamma-butyrolactone binding protein"
/product="gamma-butyrolactone binding protein"
/protein_id="CA07628.1"
/db_xref="GI:3425861"
/db_xref="GOA:086852"
/db_xref="UniProt/TREMBL:O86852"
/translation="MAKODRAIRFRTQITLDAADAVPEKQYQATITELKAVGTGK
ALYFHSKESIALGVPAOBPOVPEOPRLQELIDMGMLFCHRLRTNVAVAGRG
LSMQOAGHGDNRGPRFRKHTLKLNOAENELPHVTTTSADLVYSTRGIVY
VSQVSDIODELRHYALQKHILPALINAVPSVLAALDSEERGAFLAFLATGKD"
complement(3795..4529)
/gene="orfX"
/complement(3795..4529)
/gene="orfX"
/codon_start=1
/translation="Ketoacyl-ACP/CoA reductase"
/product="Ketoacyl-ACP/CoA reductase"
/protein_id="CA07629.1"
/db_xref="GI:3425862"
/db_xref="GOA:086853"
/db_xref="UniProt/TREMBL:O86853"
/translation="WIGSGRIGIGIALRLADGALVAHYSGSEAAARETETIRSG
GGALAIKAEIGVGDAAALYAAPDAGGEGPEPDILVNNAGSGSGITVTEES
VFDRVAVNVVAPLPIVOHGLKRLDGRILINISSAATRRAPPSIGVAMKGAVDIL
TLALRQIGERGITVNAVAPGVETDMNARRRORPEAAALAAANSVNRIGRPDIAD
VVALASDDSWITGQYVDATGCTIL"
complement(4746..6446)
/gene="orfZ"
/complement(4746..6446)
/gene="orfZ"
/codon_start=1
/translation="hypothetical protein"
/product="hypothetical protein"
/protein_id="CA07630.1"
/db_xref="GI:3451475"
/db_xref="UniProt/TREMBL:O86854"
/translation="MMRVRRGRLVGRRRRRPKASRPDGAAPARWSSPPAARORA
GGPRTADARARAYDPAAGARLPEADQAEPIPAEQACTSSAEGATMAEPAS
PGTAAEPASSPADTGQSVAVRGRTVSSAPPAKAPRMAAPDDGDRTPEPAL
PBRRAAGPAAATWGNAMVTALEGGALDAERLGRGYSHEGVDAITVPGILVLY
VRSRSRPYQVRLRTIGSDMDRFLDAAVERPGHIALDGLPSHLDLADRGVP
LIPGGLAPRCSCPDGSHGCKHAALCYOARLDDPFLILRRGRBALDALS
RRNARARAAQDRGPELPGVRAALITPAPLPAIPHPPEOPPYAPAPGS
PYPPDLHLATDAARAHALITGRDPVGLITLQDANRLAARPGSGLTRGTALTA
SLAAGRDTEALARAAVAMRGGLAGLDVLEBMDPAGFRPARPMMLAADIPAR
PMNRRLTHPRGHVQLRLGRITLWYAVESSELRBDMWRGTPDLPVALTGLGSPGDP
"

ORIGIN
Query Match 80.0%; Score 19.2; DB 1; Length 6727;
Best Local Similarity 75.0%; Pred. No. 3.9e+03;
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Cy 1 GTCTGTGCGCCGCTGACGCGAC 24
Db 2021 GTCTGTGATGCCGCTGACGCGAC 2044

RESULT 11
AP003954/c 131711 bp DNA linear HTG 21-MAR-2002
LOCUS Oryza sativa (japonica cultivar-group) chromosome 7 clone
DEFINITION OJ1381_B07, *** SEQUENCING IN PROGRESS ***.
ACCESSION AP003954
VERSION AP003954.1 GI:15021924
KEYWORDS HTG; HTGS; PHASE2.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Einhartoideae; Oryzaceae; Oryza.
1
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
clone:OJ1381_B07
Published Only in Database (2001)
2 (bases 1 to 131711)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (25-JUL-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..131711
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="7"
/clone="OJ1381_B07"

ORIGIN
Query Match 80.0%; Score 19.2; DB 2; Length 131711;
Best Local Similarity 75.0%; Pred. No. 2.5e+03;
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Cy 1 GTCTGTGCGCCGCTGACGCGAC 24
Db 95743 GTCTGTGACGCGCTGACGCGAC 95720

RESULT 12
AP004269 145772 bp DNA linear PLN 22-JUL-2004
LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,
DEFINITION PAC clone:P0048D08.
ACCESSION AP004269
VERSION AP004269.4 GI:50509267
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Einhartoideae; Oryzaceae; Oryza.
1
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
clone:P0048D08
Published Only in Database (2001)
2 (bases 1 to 145772)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (17-OCT-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Jul 22, 2004 this sequence version replaced gi:23616992.
COMMENT

```

Genes were predicted from the integrated results of the following:

GENSCAN (<http://CCR-081.mt.edu/genSCAN.html>), FGENESH (<http://www.softberry.com/>), Genemark.hmm (<http://opal.biology.gatech.edu/Genemark/>), GlimmerM ([http://www.tigr.org/tcb/glimmer/glmr\\_form.html](http://www.tigr.org/tcb/glimmer/glmr_form.html)), RiceHMM (<http://xrp.dna.affrc.go.jp/RiceHMM/>), SplizePredictor (<http://bioinformatics.laastate.edu/cgi-bin/bp.cgi>), sim4 (<http://www.tigr.org/software/glimmer/>), BLASTN and BLASTX. The genomic sequence was searched against NCBI Nonredundant Protein database, nr (<ftp://ncbi.nlm.nih.gov/blast/db>) and the cDNA sequence database at RGP or DBJ. Protein homologs of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no.

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.

The orientation of the sequence is from T7 to SP6 of the PAC clone. This sequence of P0048D08 clone has an overlap with OSJNB0041106 (DBJ: AP005176) clone at 5' end and with P0650C03 (DBJ: AP005320) at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/Genomeseq.html>.

FEATURES  
source location/Qualifiers

```
1..145772
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="7"
/clone="P0048D08"
5733..8129
/gene="P0048D08.101"
join(<5733..5870,6406..6555,6829..6905,7181..7504,
7724..7765,7865..7964,8064..>8129)
/gene="P0048D08.101"
/note="start and end point are not identified"
join(5733..5870,6406..6555,6829..6905,7181..7504,
7724..7765,7865..7964,8064..8129)
/gene="P0048D08.101"
/note="similar to Oryza sativa chromosome 1,
OSJNB0091E23.21"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD30555.1"
/db_xref="GI:50509268"
/translation="MEDGSVLPARLRPSNTVPIPTNORSTOAAHDSSEKRCV
TDESLTVGGRVSHITLIDKGVAPNPTDLSHSTGHRWTRKDAISH
TASIGARITILARFNSRGVTTTASSTNDVASLPLPLQVIELALLSPRAY
RFSHFHFKWSTYDCLVDORHSVPTSTSSGVTTTALSTSDVASLPLPLQV
ELRLGPRPAIDFSTYLYFNKVDPRPRNRLDLGTPRTDKALDPEVKDHRSG
RAGQSKSLILAEFRSDKAG"
10272..10955
/gene="P0048D08.102"
<10272..>10955
/note="P0048D08.102"
/gene="start and end point are not identified"
10272..10955
/gene="P0048D08.102"
/note="similar to Oryza sativa chromosome 7, P0048D08.3"
/codon_start=1
/product="hypothetical protein"
```

```
/protein_id="BAD30555.1"
/db_xref="GI:50509268"
/translation="MPRRDCPSRLIDQLITAKLIDINSPTDENSNDAGYKSMGR
LRHQQLPPLILPAVLSSMALPRSSHSISGPIHCSSEALAVEAKGVKSPGS
SGIPKNSFFKGDVRLTGLTALRLVMWLGAVVSDSDADGHLVITGNFP
NPPDDPFCQVRAVKDVKLPARRPAPTPANMAAPRTTAGKLPRLKMFLEKQLRKS
ANPALLAS"
14881..15381
/gene="P0048D08.103"
<14881..>15381
/gene="P0048D08.103"
/note="start and end point are not identified"
14881..15381
/gene="P0048D08.103"
/note="similar to Oryza sativa chromosome 7, P0048D08.3"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD30557.1"
/db_xref="GI:50509270"
/translation="MALPRKSSPRSSRPITIHRAVDAGRVSKPSKSGPYKSSAP
STAPVFFKGDVRLTGLTALRLVMWLGAVVSDSDADGHLVITGNFP
RDPFRVRAVKDVKLPARRPAPTPANMAAPRTTAGKLPRLKMFLEKQLRKS
EALPAS"
complement(16394..21283)
/gene="P0048D08.104"
complement(join(<16394..16504,17089..17222,17771..17894,
17988..18257,18967..19161,19242..19334,19417..19545,
19652..>21283))
/gene="P0048D08.104"
/note="start and end point are not identified"
complement(join(16394..16504,17089..17222,17771..17894,
17988..18257,18967..19161,19242..19334,19417..19545,
19652..21283))
/gene="P0048D08.104"
/note="predicted by FGENESH etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD30558.1"
/db_xref="GI:50509271"
/translation="MAGVVGSKSNRDMSDCDAFREAGRAIVALTTEEGOGHAGL
SLATLVGRYPWSPILARAILARCYLQNSRQGEVQLELAIVLAPRPHIASLIDAL
ISMDLPDEAAVRRAIVAAPEPTPALHYFVSRSASDHNDPFLERKARGRET
IRGQRARITBKGQAAASPAPASPREPPEPTVDIGADKRSMSSEERALKVSG
EMKSYCRSGMDMTSMUSDNEVFKGSGPPFCSGNITYEFAFKSHIDBEHIVGKE
FLSLVPSRISERSELLRSRWEPETDDDLAGRTKILREKELVFELEVSINLL
YIMKEFTMRVRYPAVLVSNCGSCGIGDLSSTLQELCEILKPLKLVOTRGMEHQ
KHNHDSOODSLVHTHRGCNHRKNSRQESQODSLVGIITWSEGTGLSPFCERKI
ASRPTDSSQADRPLFACILSEPLEEDMELCFMSWCPFDGPDILNINISRAIGKAL
KFSWEELKGIQGSVYFLPKAIPERDDITITPDSWIGSARVEMLLDAVDYKESL
LKTOVCLAVISITAKCLMAKLVNDPLEDALAHQNGKRPQVPLDAIIRSLMHR
RFGCDLWEIPCIISPDVARVRAILLRIFRSMDCKCDLPSXAFVWDSIRSVIDE
KAGNISAVRVESILERLHAQTFILFEFKGESLVPOTLAVPSLACICLANHFGDL
IIEKKCNVNEPMKSTFPFHSINLSVSGTLETSESELLKAVDKOSVCDFRGCG
HRIIRLYMYPHPFMYIVLRPMDNKGNIHMKULISLAELDISHIEGQSESMYTL
VSAQESWKGILIDRYSQANLRPEILFEHGRKDRHLL"
21456..21816
/gene="P0048D08.105"
join(<21456..21518,21709..>21816)
/gene="P0048D08.105"
/note="start and end point are not identified"
join(21456..21518,21709..21816)
/gene="P0048D08.105"
/note="predicted by Genemark.hmm etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD30559.1"
/db_xref="GI:50509272"
/translation="MPRRRTKVAEADEBAGBQKLLFTALTACADERKGGKSYGMBE
KGGKGGKGGEL"
complement(22480..24657)
/gene="P0048D08.106"
complement(join(<22480..22595,22769..22915,23074..23122,
23246..23371,23479..>24657))
```

On or before Oct 30, 2002 this sequence version replaced  
gi:20520661, gi:20520749, gi:20520754, gi:20520755, gi:20520688,  
gi:20520689, gi:20520815, gi:20520663, gi:20520917, gi:20520666,  
gi:20520864.

```

FEATURES
    source          location/Qualifiers
        1..290850   /organism="Streptomyces coelicolor A3(2)"
                                /mol_type="genomic DNA"
                                /strain="A3(2)"
                                /db_xref="taxon:100226"
        97..1101    /gene="SC06257"
                                /note="synonym: SCAH10.22"
                                97..1101
        /gene="SC06257"
                                /note="SCAH10.22, probable ABC transport system sugar
                                binding lipoprotein, len: 334 aa; similar to TR:CA841563
                                (EMBL:U049727) Streptomyces coelicolor putative secreted
                                solute binding protein, 337 aa; fasta scores: opt: 1003
                                z-score: 1130.9 E(): 0; 49.8% identity in 325 aa overlap,
                                to SW:RBSB_BACSU (EMBL:Z92953) Bacillus subtilis
                                D-ribose-binding protein precursor RBSB, 305 aa; fasta
                                scores: opt: 206 z-score: 238.7 E(): 6.3e-06; 23.0%
                                identity in 235 aa overlap and to TR:CA841563 (EMBL:SC9B1)
                                Streptomyces coelicolor SC9B1, 17c, 337 aa; fasta scores:
                                opt: 1003 z-score: 1045.2 E(): 0; 49.8% identity in 325 aa
                                overlap. Contains match to Pfam entry PF00532
                                Peripla_BP_1like, Periplasmic binding proteins and Lacti
                                family and a match situated in the correct position to
                                Prosite entry PS00013 Prokaryotic membrane lipoprotein
                                lipid attachment site. Contains also possible N-terminal
                                region signal peptide sequence"
                                /codon_start=1
                                /transl_table=11
                                /product="putative ABC transport system sugar binding
                                lipoprotein"
                                /protein_id="CA261076.1"
                                /db_xref="GI:6273662"
                                /db_xref="UniProt/TR:EMBL:O9RKT5"
                                /translation="MARFRTVGIIVLAGALSVLSIAGCSSYTGKRAEDARKAASAGRA
                                AVLPRTFAMITHSGDGFPMIDVVGAGQAQAKVDINFLYSHDAIAQOQAQALVNAQA
                                IDKKVDGIITVLAKPDAMKSAALRAHQAIPVITVNSGESSEKFGALFVGGDETTA
                                GEAVGEELNERRGRQACVLAHEGQANRBSQSCGVEETDPGKQKRLVYNQTSMPDQNS
                                ALEKIKQDTSVDAVVTLGAPVADTVAKAQAGSKSEIDTFPDNAAVAAGADGTGGA
                                FAVDQGPYLQGYEAVDILMDVKNYNAVLTGGGPVLTGPPIITXDAAALADLYTKRGTR
                                "
        133..165     /gene="SC06257"
                                /note="PS00013 Prokaryotic membrane lipoprotein lipid
                                attachment site"
                                241..966
        /gene="SC06257"
                                /note="Pfam match to entry PF00532 Peripla_BP_1like,
                                Periplasmic binding proteins and Lacti family., score
                                13.80, E-value 2.1e-06"
                                1098..2138
        /gene="SC06258"
                                /note="synonym: SCAH10.23"
                                1098..2138
        /gene="SC06258"
                                /note="SCAH10.23, probable ABC transport system sugar
                                permease, len: 346 aa; similar to TR:O68120
                                (EMBL:AF010496) Rhodococcus capsulatus ribose transport
                                system permease protein RbsC; fasta scores: opt: 445
                                z-score: 508.0 E(): 6.4e-21; 30.5% identity in 351 aa
                                overlap and to SW:RBSB_ECOLI (EMBL:U10328) Escherichia
                                coli ribose transport system permease protein RbsC, len:
                                321 aa; fasta scores: opt: 286 z-score: 329.9 E():
                                5.3e-11; 30.1% identity in 329 aa overlap. Contains
                                possible hydrophobic membrane spanning regions"
                                /codon_start=1
                                /transl_table=11
                                /product="putative ABC transport system sugar permease"

```

/protein\_id="CAB60177.1"  
/db\_xref="GI:6273663"  
/db\_xref="GOA:Q9RKT4"  
/db\_xref="UniProt/TREMBL:Q9RKT4"  
/translation="MNATOVKADDERIIQTSRLRLGLGRPEIGSVGAIAVEFPAF  
ADSPHAASTLVLAFTIGMAVPVALLMGEPFDLSAGMTWSLVSMSYOM  
TANWVGVSILVTLATLGAIPGFWLPTPKLPSFIIITCTGLMTGLNGLKIDGT  
VSTKSIADMEGPPSAQDVPAFTITIGVGFKTIIIMLALVAVSMILITRAQWIF  
AVGNKDAARAVGVPACTKIGLVGVGFGAMIISGQHLFSYDVVQSGEGVNELIYF  
IAAVIGCLITIGYGSVAVGAFIPGWTSGKIVFAEWBNDFWFKFLGAMILLATLL  
NANVVRKRAETK"  
2124. 2127  
/gene="SC06258"  
2135. .2926  
/gene="SC06259"  
/note="synonym: SCAH10.24"  
2135. .2926  
/gene="SC06259"  
/note="SCAH10.24, probable ABC sugar transport ATP binding  
protein, len: 263 aa; similar to TR:068121 (EMBL:AF010496)  
Rhodobacter capsulatus ribose transport ATP-binding  
protein, 305 aa; fasta scores: opt: 612 z-score: 706.1  
E(): 5.9e-32; 45.6% identity in 217 aa overlap, to  
SW:RBSA.ECOLI (EMBL:M1169) Escherichia coli; ribose  
transport ATP-binding protein RbsA, 501 aa; fasta scores:  
opt: 509 z-score: 585.6 E(): 3e-25; 31.5% identity in 254  
aa overlap and to TR:CA15787 (EMBL:SC7B7) Streptomyces  
coelicolor ST787.07, 260 aa; fasta scores: opt: 592  
z-score: 521.2 E(): 1e-23; 43.1% identity in 246 aa  
overlap. Contains Pfam match to entry PF00005 ABC tran.  
ABC transporter and Prosite PS00017 ATP/GTP-binding site  
motif A (P-loop) and PS00211 ABC transporters family  
signature"  
/codon\_start=1  
/transl\_table=11  
/product="probable ABC sugar transport ATP binding  
protein"  
/protein\_id="CAB60178.1"  
/db\_xref="GI:6273664"  
/db\_xref="GOA:Q9RKT3"  
/db\_xref="UniProt/TREMBL:Q9RKT3"  
/translation="MTBMBDRPRLVLSGVSNNYGNVAVLGVSLVFNAGRTCVLGD  
NAGKSTIKITAGLHNDGTLISLDGETRISLRRLADRIATVYDGLAVPLMPY  
WNPFLLGSPRKGVAFFRMDVHMRRTTHALLMGIDLDDVDPIGLTSGEGRQCV  
AIARAVYGAKTVLDEPTALGVQSGVAVVAADQGLGVLLITHNPHAYLGV  
DRFVLKRGITWVAQTRDEVTIDELTQMGAGTETDIDRHLERNG"  
2240. 2815  
/gene="SC06259"  
/note="Pfam match to entry PF00005 ABC tran, ABC  
transporter, score 168.70, E-value 9.6e-47"  
2261. .2284  
/gene="SC06259"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
2588. 2632  
/gene="SC06259"  
/note="PS00211 ABC transporters family signature"  
3042. .4190  
/gene="SC06260"  
/note="synonym: SCAH10.25"  
3042. .4190  
/gene="SC06260"  
/note="SCAH10.25, possible sugar kinase, len: 382 aa;  
similar to SW:GIK.STRCO (EMBL:X65932) Streptomyces  
coelicolor glucokinase (EC 2.7.1.2) (glucose kinase) GIK,  
317 aa; fasta scores: opt: 1022 z-score: 1147.9 E(): 0;  
46.6% identity in 311 aa overlap and to TR:CAB51974  
EMBL:SC6B10 Streptomyces coelicolor SC6B10.20c, 317 aa;  
fasta scores: opt: 1022 z-score: 1035.4 E(): 0; 46.6%  
identity in 311 aa overlap. Contains a match to Pfam entry  
PF00480 ROK. ROK family and Prosite entry PS01125 ROK  
family signature"  
/codon\_start=1  
/transl\_table=11

/product="putative sugar kinase"  
/protein\_id="CAB60179.1"  
/db\_xref="GI:6273665"  
/db\_xref="GOA:Q9RKT2"  
/db\_xref="UniProt/TREMBL:Q9RKT2"  
/translation="MSTYRDPFAIGSRAPVLTGTRRRSHLTAPRVETGIDIG  
GTFKMAVVAADGNIILEKTRTEPDKRSKRVVEDTIVELDLSDDHVAVGIGAA  
GMVDADNRRIYFAHLSKRNBPRLDRLAGLAIVYVYDNDANTAAAMEMRGARGSD  
HLVWITIGGAILLEDGVYKRYVAGFPMQVPGGHRPCPCNKGWBOYSSG  
NALVREARELAAADSPVAYGIIERXKSIDITGPMITTELRBDNACVELLDIIGW  
LGVGIANLAAALDPSCFVIGGVSAADDLIGPARDAFKROLTGRGYRPEARIYVRAOL  
GPEAGVGAADLSRLVARRPRARREVERERYERYAARESRSL"  
3159. 3704  
/gene="SC06260"  
/note="Pfam match to entry PF00480 ROK, ROK family, score  
215.50, E-value 5.4e-63"  
3540. .3623  
/gene="SC06260"  
/note="PS01125 ROK family signature"  
4244. .4247  
4259. .4768  
/gene="SC06261"  
Query Match 80.0%; Score 19.2; DB 1; Length 290850;  
Best Local Similarity 75.0%; Pred. No. 2.2e+03;  
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCCGTGTGCCGTSACSCGSAC 24  
Db 12986 GTCCGTAGGCCCGGTGACCCGCAC 12963  
RESULT 14  
BC032284/c 1851 bp mRNA linear ROD 29-JUN-2004  
LOCUS BC032284  
DEFINITION Mus musculus porcupine homolog (Drosophilal), transcript variant  
Mporc-c, mRNA (CDNA clone MGC:40733 IMAGE:5362485), complete cds.  
ACCESSION BC032284  
VERSION BC032284.1 GI:21595487  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1851)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Sherman,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Wax,S.I., Wang,J., Heideh,F.,  
Dietchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Schaeer,T.E., Brownstein,M.J., Ubedi,T.B., Toohily,K.I.,S.,  
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.U.,  
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.U., Malek,U.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.U., Hulyk,S.W.,  
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,  
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalins,D.E.,  
Scherer,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
JOURNAL PUBLISHED 12477932  
2 (bases 1 to 1851)  
Strausberg,R.  
REFERENCE Direct Submission  
TITLE Submitted (06-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,



REMARK  
COMMENT

USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Web site: <http://www.nisc.nih.gov/>

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAK Plate: 65 Row: d Column: 4  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 22094150.  
Location/Qualifiers

FEATURES  
source

1. 1851  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="MGC:40733 IMAGE:5362485"  
/cissue\_type="Eye, retina, mouse strain C57Bl/6"  
/clone\_lib="NIH MGC\_94"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. 1851  
/gene="Porcn"  
/note="synonyms: Mporc-a, Mporc-b, Mporc-c, Mporc-d, Ppn,  
Mporc, porc"  
/db\_xref="LOCUSID:53627"  
/db\_xref="GI:21595488"  
/db\_xref="MGI:1890212"  
144. 1511  
/gene="Porcn"  
/product="porcupine homolog"  
/protein\_id="AA32284.1"  
/db\_xref="GI:21595488"  
/db\_xref="LOCUSID:53627"  
/db\_xref="MGI:1890212"  
/translation="MATESROEPFOQLQGLPTVOGQIDQIWLITICPACLLMR  
LGLPSYKHAISTVAGGFFSLYHFFQILMNVVVLISLCYLVPLCRSHSRGVLSVT  
ILYILMGEMHMDVTVMHMRGQOMVAMKAVSLGPDLDGEVGAVPSPVEFWGYL  
FVGTIVGPMWISFHSYLQAVQGRPLSRMLKVARSLALCLVSTCVGYLPYF  
IPLDGRLLRKGTWMLRAYESAVSFHSYFVGLSEATATLAGAGFTEKDLHEM  
DLTVSRPLNVELPRSNVEVTSKMLPMSYMLNVCVPLKALGTEFAVLVYASALL  
HGSEFHALAVLISLAFITVEHVLRKLAQILISACILSKRCLPDCSHRHRLGIVRAL  
NLRGALAIFFHLSYGLSPVDVDDTTEQSGMAVYHKMSLSMASHWTFCWTF  
YRLIG"

ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 1851;  
Best Local Similarity 70.8%; Pred. No. 6.6e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGCTGCGCGCTGACGCGAC 24  
|||||:|||||:|||||:|||||:  
Db 90 GTCTGCGCGCGCGGACGCGAC 67

RESULT 15  
AB036747/c  
LOCUS AB036747 1886 bp mRNA linear ROD 20-JUL-2000

DEFINITION  
AB036747  
VERSION  
AB036747.1 GI:6714507  
KEYWORDS  
porcupine-A.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.  
1 (sites)  
Tanaka, K., Okabayashi, K., Asashima, M., Perrimon, N. and Kadowaki, T.  
The evolutionarily conserved porcupine gene family is involved in  
the processing of the Wnt family  
Eur. J. Biochem. 267 (13), 4300-4311 (2000)

REFERENCE  
AUTHORS  
MEDLINE  
PUBMED  
10866835  
2 (bases 1 to 1886)  
Kadowaki, T. and Tanaka, K.  
Direct Submission  
Submitted (08-JUN-2000) Tatsuhiko Kadowaki, Nagoya University,  
Graduate Program for Regulation of Biological Signals; Chikusa,  
Nagoya, Aichi 464-8601, Japan  
(E-mail: [emi@uagrl.agr.nagoya-u.ac.jp](mailto:emi@uagrl.agr.nagoya-u.ac.jp), Tel: 81-52-789-5237)

FEATURES  
source

1. 1886  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
1. 1886  
/gene="Mporc-a"  
174. 1526  
/gene="Mporc-a"  
/codon\_start=1  
/product="porcupine-A"  
/protein\_id="BA89468.1"  
/db\_xref="GI:6714508"  
/translation="MATESROEPFOQLQGLPTVOGQIDQIWLITICPACLLMR  
LGLPSYKHAISTVAGGFFSLYHFFQILMNVVVLISLCYLVPLCRSHSRGVLSVT  
ILYILMGEMHMDVTVMHMRGQOMVAMKAVSLGPDLDGEVGAVPSPVEFWGYL  
FVGTIVGPMWISFHSYLQAVQGRPLSRMLKVARSLALCLVSTCVGYLPYF  
IPLDGRLLRKGTWMLRAYESAVSFHSYFVGLSEATATLAGAGFTEKDLHEM  
DLTVSRPLNVELPRSNVEVTSKMLPMSYMLNVCVPLKALGTEFAVLVYASALL  
HGSEFHALAVLISLAFITVEHVLRKLAQILISACILSKRCLPDCSHRHRLGIVRAL  
NLRGALAIFFHLSYGLSPVDVDDTTEQSGMAVYHKMSLSMASHWTFCWTFYRLIG"

ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 1886;  
Best Local Similarity 70.8%; Pred. No. 6.6e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGCTGCGCGCTGACGCGAC 24  
|||||:|||||:|||||:|||||:  
Db 120 GTCTGCGCGCGCGGACGCGAC 97

RESULT 16  
AB036748/c  
LOCUS AB036748 1901 bp mRNA linear ROD 20-JUL-2000  
DEFINITION  
Mus musculus Mporc-c mRNA for porcupine-C, complete cds.  
AB036748  
VERSION  
AB036748.1 GI:6714509  
KEYWORDS  
porcupine-C.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.  
1 (sites)  
Tanaka, K., Okabayashi, K., Asashima, M., Perrimon, N. and Kadowaki, T.  
The evolutionarily conserved porcupine gene family is involved in  
the processing of the Wnt family  
Eur. J. Biochem. 267 (13), 4300-4311 (2000)

REFERENCE  
AUTHORS  
MEDLINE  
PUBMED  
10866835

## REFERENCE

2 (bases 1 to 1901)  
 Kadowaki,T. and Tanaka,K.  
 Direct Submission  
 Submitted (08-JAN-2000) Tatsuhiko Kadowaki, Nagoya University,  
 Graduate Program for Regulation of Biological Signals, Chikusa,  
 Nagoya, Aichi 464-8601, Japan  
 (E-mail:eml@nagrl.agr.nagoya-u.ac.jp, Tel:81-52-789-5237)

## FEATURES

source  
 location/Qualifiers

1..1901  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 1..1901  
 /gene="Mporc-c"  
 174..1541  
 /gene="Mporc-c"  
 /product="porcupine-C"  
 /protein\_id="BAA89469.1"

## CDS

gene

## ORIGIN

## Query Match

Best Local Similarity 78.3%; Score 18.8; DB 10; Length 1901;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGGCCGCGTSACGCGSAC 24  
 120 GTCTGTGGGCGCGGAGCGGCAC 97

## Db

RESULT 17  
 AB036746/c 1904 bp mRNA linear ROD 20-JUL-2000  
 LOCUS Mus musculus Mporc-b mRNA for porcupine-B, complete cds.  
 DEFINITION AB036746  
 AB036746.1 GI:6714505  
 VERSION porcupine-B.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus

## REFERENCE

1 (sites)  
 Tanaka,K., Okabayashi,K., Asashima,M., Perrimon,N. and Kadowaki,T.  
 The evolutionarily conserved porcupine gene family is involved in  
 the processing of the Wnt family  
 Eur. J. Biochem. 267 (13), 4300-4311 (2000)

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

source  
 location/Qualifiers

1..1904  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 1..1904  
 /gene="Mporc-b"  
 174..1544

## CDS

## ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 1904;  
 Best Local Similarity 70.8%; Pred. No. 6.5e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGGCCGCGTSACGCGSAC 24  
 120 GTCTGTGGGCGCGGAGCGGCAC 97

## Db

RESULT 18  
 AB036749/c 1919 bp mRNA linear ROD 20-JUL-2000  
 LOCUS Mus musculus Mporc-d mRNA for porcupine-D, complete cds.  
 DEFINITION AB036749  
 AB036749.1 GI:6714511  
 VERSION porcupine-D.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus

## REFERENCE

1 (sites)  
 Tanaka,K., Okabayashi,K., Asashima,M., Perrimon,N. and Kadowaki,T.  
 The evolutionarily conserved porcupine gene family is involved in  
 the processing of the Wnt family  
 Eur. J. Biochem. 267 (13), 4300-4311 (2000)

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

source  
 location/Qualifiers

1..1919  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 1..1919  
 /gene="Mporc-d"  
 174..1559  
 /gene="Mporc-d"  
 /product="porcupine-D"  
 /protein\_id="BAA89470.1"  
 /db\_xref="GI:6714512"  
 /translation="MATESROEFFQOLLQGLPTVOOGLDQIWLTLTICFACRLMR  
 LGLPSYLKHAFTVAGGFSPLVHFMOLMMWVVLISLCYLVLFCRSHSHRGVFLSVT  
 ILIYLKGMENMDVTYVHKRKGOMIVAKKAVSLGPDLDGEGAVPSVPEFMGYLY  
 FVGITVFGPMISFHSYLOAVQGRPLSRKWKVARSIALALCLVLTCTGCPYLPFYF  
 IPLDGRLLRNKRKARWMLRAYEAVSFHSNPFVGLSEATATLGAAGFTTEKHLEW  
 MDLTVSRPLNVELPRSMVEVVTSMNPKSVWLNWVFKNALRLGTFSAVLVYASAL  
 LHGFSFHLAIVILSLAFITYVEHVLKRLAQIISACILSKRCIPDCSHHRLGLGVRL  
 LNLFGALAIPLHLSYGLSLFDVVDVDTTEEGYGMAYTVHKMSLSMAHMTVFCGWI  
 FRLIG"

## ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 1919;  
 Best Local Similarity 70.8%; Pred. No. 6.5e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCCGTGCGCGCGTSGACGCGSAC 24  
 |||||:||||:||||:||||:||||:  
 Db 120 GTCCGTGCGCGCGCGCGACGCGCAC 97

RESULT 19  
 SV117268 8992 bp DNA linear ECT 02-AUG-2002  
 LOCUS Streptomyces viridochromogenes phbB gene, phcB gene, ORF1 and  
 DEFINITION ORF1.  
 VERSION Y17268  
 KEYWORDS ORF1; ORF1; phosphinothricin tripeptide synthetase B;  
 phosphinothricin tripeptide synthetase C; phbB gene; phcB gene.  
 SOURCE Streptomyces viridochromogenes  
 ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycetaceae; Streptomyces.

REFERENCE 1  
 AUTHORS Schwartz, D., Grammel, N., Reckienwald, J., Keller, U. and Wohleben, W.  
 TITLE Isolation and characterization of the peptide synthetase genes phbB  
 and phcB from the phosphinothricin tripeptide producer Streptomyces  
 viridochromogenes  
 JOURNAL unpublished  
 REFERENCE 2  
 AUTHORS Schwartz, D.  
 TITLE Direct Submision  
 JOURNAL Submitted (07-MAY-1998) D. Schwartz, Universitaet Tuebingen,  
 Lehrstuhl Mikrobiologie/Biotechnologie, Eberhard-Karls-Universitaet  
 Tuebingen, Auf der Morgenstelle 28, D-72076 Tuebingen, FRG  
 revised by [3]  
 REFERENCE 3  
 AUTHORS Schwartz, D.  
 TITLE Direct Submision  
 JOURNAL Submitted (02-AUG-2002) D. Schwartz, Universitaet Tuebingen,  
 Lehrstuhl Mikrobiologie/Biotechnologie, Eberhard-Karls-Universitaet  
 Tuebingen, Auf der Morgenstelle 28, D-72076 Tuebingen, FRG  
 On Aug 4, 2002 this sequence version replaced gi:8250615.  
 COMMENT location/Qualifiers  
 FEATURES source  
 1..8992  
 /organism="Streptomyces viridochromogenes"  
 /mol\_type="genomic DNA"  
 /strain="Tue494"  
 /db\_xref="taxon:1938"  
 209..418  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAB93682.1"  
 /db\_xref="GI:8250616"  
 /db\_xref="UniProt/TREMBL:Q9KMY9"  
 /translation="MIDGATRFRLVYVNHGQYVSWPEDRPAGMRPHGTRAC  
 LEVIRHMDRPLSLKSTFEEGA"  
 415..3984  
 /gene="phbB"  
 415..3984  
 /gene="phcB"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphinothricin tripeptide synthetase B, PhbB"  
 /protein\_id="CAB93683.2"  
 /db\_xref="GI:22095156"  
 /db\_xref="GOA:Q9KMY8"  
 /db\_xref="UniProt/TREMBL:Q9KMY8"  
 /translation="MTAPQTDVVYTGRIATWAGLRLPEIGIDNNVFRIGASSVMAV  
 RAAARIRALDTPLRLDVESPPALAKIRASRTAPASPPTAPADSTATAP  
 LTFQOBPMWLFDROMPGNATYTIHFALRHGHLDLGYLDRCVRRVVARHVALRTSPS

VDNRPAQOVIDRAHPIPGESDRLALPAERPVAAARIAAREQAPELSTGPMRLRVY  
 LHSIDTQRIULLTTPHIVTDWSDDLVRELNHLYRATGTGIAQALPPLPQVDMKA  
 RORAEIQTAEELLDWNRHLAGVPLPRLPAPRAVVRHRRGLLFLPIPESTVR  
 LEGIAKDEGTTPYAVILAGPALLHRTIGQDDLVGSPVGRTEETEGEGLVFNVA  
 ARCDVGRSPLEIVRTRRTVRESPAROELPHRLVVELIAPVRSAYPLVQVMA  
 LQNPDDGRRPGRGPARRESGDGTSANFDLITLPTGASGRCGMEWEDSDLPDE  
 RVNELGPDLVLDALDRDLDPALPLPDEPARDMVQAMNDTALVRCGPADSL  
 PALLSAQHRTPDVALRTDDGALTYRQLHRLRDRARLISGLAPESVAVACSR  
 SPEWVVALILVAKGCAVLPIDPDPAERTAYILRDSGARVLLTLHRTNLPDADGT  
 TVVTLDEPDSGMODTTSALPGIAPQOLAVLYTSGSGRPGKVLNHEGVCNRIEM  
 GMAPRPGCTIVLOKTIHPDVSVMEMFTLATGATLVILAPDGHDDPOLVRLVE  
 EGTVDVFPVSMALAPLDVICALPBGHSLRVFCSGEGLSPGLDRILPARIPIHEINL  
 YGPTAEALIVTHWRCRGPETVPIGRITANARCYLDAELNPPVPGVGEMLGCVV  
 ARGHGRGADLTAERFLPDPYGPGRSMYRSGDLARMRDVLLEYLGRDQVXLRGOR  
 LEIIEGATLLAGHAEVADVVDVYRGTPQDORLIVA YVRPARGDEQLRTTLRELA  
 RLAPVNRSSVYTLDRVPLPSPGTDKALPDPAAGOPRSGRAAGTAPERLTIAGI  
 MARLIGAEVGDNDPPEIGHSILARBMTRGATLSTARGNDLPVSLAEHEVLRPFLA  
 VTRQATDSALATERLLAELEBALADTELEALPDDEGPDGSGE"  
 3977..5638  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAB93276.1"  
 /db\_xref="GI:22095157"  
 /db\_xref="UniProt/TREMBL:Q9KJL4"  
 /translation="MSESTTGASDSTSDGTRGPRSGARPARADGMAVAGVLIG  
 ALICENVIIIVFKTVARFGGSAVVVVGAWMLRRRLSPSLPVVSISSGVLISA  
 AVCSGIAANILRGAGPPPVVALAIVILANAVGLVGLMLAARLRSREILPPTLRPAI  
 LMESLSRKRREVSARPLMSALASVVVALVASTDGTGIALHPLGPAFVGLASPMIL  
 VGALAGSRSLAMWPFVGVGSAAVFWARDDADDTYDLSHGVALVIGIYIGVSWT  
 LVRSVSLRRPGCATVPDGRAGALCVLLAGALTLCHWYGRGLGVALTLVMT  
 LFLALFEVRGAETGIALSPALFELGILRLMSDAILVSTVAVAGAMALYYT  
 YVARVSPITAVSAPSPRIVSTOALGCMGIVAGLPLMAVLVLSVAGGQBPPLAR  
 APFVSAHYSLDSDAQOCTPLAALPAPFLITRAPVYSALGELLPLPDVTLANS  
 LGGAAMLTVRKAPHLKETVGTVASGLVIGSVASAVIVARAFG"  
 5711..8971  
 /gene="phcB"  
 5711..8971  
 /gene="phcB"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphinothricin tripeptide synthetase C, PhcC"  
 /protein\_id="CAB93684.1"  
 /db\_xref="GI:8250618"  
 /db\_xref="GOA:Q9KMY7"  
 /db\_xref="UniProt/TREMBL:Q9KMY7"  
 /translation="MEDLQTRIALSLSKQALFESRLRAAAAPQDPALIRPDDGCP  
 VPISFAQRIIMFLDLEPRGVNVNSASLIGRPVTEAVRDALGALTRHBYLRTVP  
 PADGSRHOIADSLTPPLTETDRLALPDSARAALRLCAEDKORFPDLSGFLAC  
 LLIRIRDDALLFLTFHHVTFPDGMSIGLIRDLTLALHAAGTGDAGLPPLIOYADP  
 ADMQRRLDEKRLGELGTYRERIRKAPPIYIDPDRPRPAVATTBEARRRRPALRL  
 TVALRDLAARSVAITPFTMLTVFPAALHARSRSRDVITGVNARARLELDLGFVA  
 NTLAMVRIEPRGMSFGDLAQVQVVEALATLELAQWCDGFLGVGAGLADDPDVT  
 APFWEDGRDASELDLTLPERADTHETPDSKFLVTLVLTBERSTYTFEYNTALFEP  
 VTIDRLGERLALLARSVAADPGWELALPVLTDEIRHLEKVNAPVADDERHRTLLG  
 VLDSARRHDDVAARADROLTYRLEDEANLAHLHLLITQSLAGRPQVALDGTAD  
 AITAVAVAKAGAVLPIDPEYPAERLHILRSRGAITLLTQSLAGRPQVALDGTAD  
 LDDATATVALADGADPDPGLPIADRIAYITPISGSGVGVKGVNVPRAQSLRYSB  
 QFAQTPDSRVLRPASBSFDVLSLELMTLPAGATVLVLEPPALVPGEDLARLIRERY  
 STVILSPASLSTLTAGELPGLRTVVMAGAAATLELAQWCDGFLGVGAGLADDPDVT  
 TIARCAPDVPGLPAGVAVVLDLDTLAVPFGSGELFLGVLGAGLADDPDVT  
 ADRFLPDPSGTEPGRVLYRTGDVVRWKADELFLFGRTHOVLRGRIILGELRIL  
 EDHPGRTAVVLYRGESDRRLAGYVRAAGKRRPPTAAGICRWLRDLPRYMWPELT  
 VLALPTSPNGKLDREALPDLAQSGTDGAGRPLDLPDVERLSGIWQVYGLAPPS  
 ADNFEVGSNSLSATRIARVNOAFVRLPVRSLFVEPTLSGLARSVABARELP"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 1; Length 8992;  
 Best Local Similarity 70.8%; Pred. No. 5.2e+03;  
 Matches 17; Conservative 2; Indels 0; Gaps 0;

Qy 1 GTCCGTGCGCGCGTSGACGCGSAC 24

Db 3420 GTCCTGCGCGCGGTGCCCGGAC 3397

|||||:|||||:|||||

RESULT 20  
AE012334/C

LOCUS 11004 bp DNA linear BCT 23-MAY-2002  
DEFINITION Xanthomonas campestris pv. campestris str. ATCC 33913, section 242

ACCESSION AE012334 AE008922

VERSION AE012334.1 GI:21113402

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Xanthomonas campestris pv. campestris str. ATCC 33913  
Xanthomonas campestris pv. campestris str. ATCC 33913  
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
Xanthomonadaceae; Xanthomonas.

1 (bases 1 to 11004)  
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Camargo, L.E.A., Camarotte, G., Camnava, F., Cardozo, J.,

Chambers, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,

Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,

Ferreira, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,

Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melandri, J.,

Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,

Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.H.A.,

Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinoia, L.A.F.,

Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezsa, R.I.D., Trindade dos

Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

Kitejima, J.P.

Comparison of the genomes of two Xanthomonas pathogens with

differing host specificities

Nature 417 (6887), 459-463 (2002)

22022145

2 (bases 1 to 11004)

da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Camargo, L.E.A., Camarotte, G., Camnava, F., Cardozo, J.,

Chambers, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,

Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,

Ferreira, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,

Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melandri, J.,

Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,

Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.H.A.,

Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinoia, L.A.F.,

Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezsa, R.I.D., Trindade dos

Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

Kitejima, J.P.

Direct Submission

Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de

Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,

Brazil

Location/Qualifiers

1. 11004

/organism="Xanthomonas campestris pv. campestris str. ATCC

33913"

/mol\_type="genomic DNA"

/strain="ATCC 33913"

/db\_xref="ATCC:33913"

/db\_xref="taxon:190485"

/note="pathovar: campestris"

/complement(119 . 607)

/gene="XCC2264"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/genemark"

/codon\_start=1

/transl\_table=1

/product="general stress protein"

/protein\_id="AA04543.1"

/db\_xref="GI:21113403"

/translation="MADPKELQDKFMRKALSDRTVMGLDGVDEGHARPMWQIEGDS

GGPIWFTTSKONALIAMLGGRRITGFPSSKGHDLPSISGSLREDPDPAVDRMLNP

YVAAMVYRGKDDPKLALLRLDADHAQIWLNGSSILAGIKVLLGVDPKDDYQDKADV

LR"

1014 . 2219

/gene="XCC2265"

1014 . 2219

/gene="XCC2265"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/genemark"

/codon\_start=1

/transl\_table=1

/product="pectinesterase"

/protein\_id="AA04544.1"

/db\_xref="GI:21113404"

/translation="MSASHRTTCLATLALSLIDNAALVTGTSRQLTSSNGTYT

ISKLAAGAPITALTVDNMPRTAGVALTADFAAAGSTRYRTQGAIPAAGAGS

TRRYSVKGATVELVCPGAPPLTIFGSAATPDVIRYNNANPTPKPGTASHFC

ASNAATVGSNATYTVASGQARHLRPNNVGTYADNNQSAFLAIVRGDOAS

FEDLVVGNDTLISATNANVIRAFKSTTIGVDYIFGSGTGFNDALITSAGA

RLGSGRGYIFAPTRPGSGGFLAINSRVAGSGSPDNDTYLGRADVEGSGTAV

NGTSPNQVITRDSLTGSHIRKTPMNASTASBPSCSNCTNSANRFFSNVTAVLV

PTDGRGRV"

2962 . 4632

/gene="pgja1"

/note="synonym: XCC2266"

2962 . 4632

/gene="pgja1"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/genemark"

/codon\_start=1

/transl\_table=1

/product="polygalacturonase"

/protein\_id="AA04545.1"

/db\_xref="GI:21113405"

/translation="MOPHPRLPALAPSLIALGLATPOASAPRPHGLVDATASIST

SMGVQOPALPSQYCTTLKALVAVASLADDDPQSRDTRARLQALIDPCAGSA

VRILVGGSGESGLSPGLTKSGVTLIDKVTLFGSRNPDYNGTCTATTDDA

KSCNPRLHVDATASAIIVGAKIDGSGTTLTAGPNKASWMDLAVLTAKGSGV

PRILQIDSDTLFTYDITLNSPHEHTTVDVGLTAMGKILAPSLVSRPGYCHA

ASTGVNPHASCFPEPTAKTNGDPCQSKVLLAYSYIGDGDGVAIKAAKRSI

ASENMLFAYNVQFYTHGSPGSESTDSGRHIAVGLSIDGNANDVQDRPYSAGRI

KSDSRGQYDITFENICMGVARPIVPAIVANVATPQDLPRPSGAILTNVSLGS

KAFGGELSPFTGIDATTLTILSLNVVDGGPVAPOHFGGSPSNPGAVHFTFK

GAPVSFPDLTBSANDVOLQGRPGAPALQCDDAFLAYHSLPDSPI"

/complement(4695 . 5384)

/gene="XCC2267"

complement(4695 . 5384)

/gene="XCC2267"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/genemark"

/codon\_start=1

/transl\_table=1

/product="conserved hypothetical protein"

/protein\_id="AA04546.1"

/db\_xref="GI:21113406"

/translation="MSQSALDSFLDKMRKMPWSVAEPPIPPORRLAAMFALLQ

EMEDIMNIGDPLPADAKLAWQOELDMWSQRHPLGRVLEVPAPMQLQTLPA

MOARAPDDIQAOMARLOAFVAIVVAVFAPRTSSSDATAVVOMLDRORVAG

SAAPGIVGAPMWRQULRAMPRKAAALAPRHWRSRLARLRLRLAAGTAYAPMQL

WHSWRASGAE"

complement(5413 . 6078)

/gene="cbdz"

/note="synonym: XCC2268"

complement(5413 . 6078)

/gene="cbdz"

/note="identified by sequence similarity; putative; ORF

RESULT 21  
AF506520

13513 bp DNA linear BCT 16-DEC-2003

AF506520

Query Match

Best Local Similarity 70.8%; Pred. No. 5, 1e+03;  
Matches 17, Conservative 5, Mismatches 2, Indels 0, Gaps 0;

1 GTCCTGATGCGCCGTCGACGCGSAC 24  
|||||:|||||:|||||:|||||:|||||:  
Db 1700 GTCCCTGGTTCGCCGTGACACGAC 1677

gene  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoglycolate phosphatase"  
/protein\_id="AA041547.1"  
/db\_xref="GI:21133407"  
/translation="MRRPRAVLPDLGDLNADPMATVYANMISERGLPCITLAAOLR  
PVYKSGRAMLAFAHDLDAAREALVPEFLKRRYBALGTQAQIPDGYEWLQRLBOA  
GTCVGYITNKRBYLAQLILPQLQOQRCAVILGGDTLAERKPHPLPLVLADRIGVAA  
TQCCYGVGDDEBDIILAAARAGMPSVAALLMWYRLGDDPLSWQADVLVEQPLMEPAAAM  
PQ"

complement (6191) . 6910)  
/gene="ubig"  
/note="synonym: XCC2269"  
complement (6191) . 6910)  
/gene="ubig"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="3-demethylubiquinone-9-3-methyltransferase"  
/protein\_id="AA041548.1"  
/db\_xref="GI:21133408"  
/translation="MNRPDOSTSGNFHOTELDKRPAALANRWADADGPKRIHALNPPR  
LDYAAARVALPGARVLDVGGCGGLISEMNAALCAVTRIDIAPELVKARHLSSESSY  
QVDRPVSQVEDLAARQPSGPDVATCMELLEVHPPLAIRACASILKRGGLFLSTLNA  
RTPAFVALVAGAEYIARILPFGKTHHYKDFIKPSBELAAMRLTAELQLODYSGLMEYEP  
RNRRLSSRTENVNLIACVAP"

complement (6939) . 8288)  
/gene="XCC2270"  
complement (6939) . 8288)  
/gene="XCC2270"  
complement (6939) . 8288)  
/gene="XCC2270"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA041549.1"  
/db\_xref="GI:21133409"  
/translation="WMDPLMTAAPPEPCDILTEAGVVPPIEBPAVVLDEHAAVANSNGVIL  
VAVLPADARVAPARVAPTRSPDALMGLVNAHTNPMITLBRVADDLPLMWLLOOH  
IWPBEAVIIGPEFVADGTTLAIEMLRGGTCVAMENFFAVQAVYKQHGFRALVGA  
VILDPYLAHTEETAQEVADSVAAQYQORPLADLDELGVNDLIVHNTQULTEAHLIC  
AERGVSVVHCPSNLIKLAGFCPCALORASVNLAIIGSGSNDDIMDFEENRTAAL  
LAKVANDATLDAATTIRATLIGGAALIGRDRIGSIEVGKADIVCVDALETOB  
LHNVLSQLIYAAGRHQYTDVWIAKPKLVQRELDIMDTALVANAQRDRIRIVRA"

complement (8294) . 9139)  
/gene="XCC2271"  
complement (8294) . 9139)  
/gene="XCC2271"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA041550.1"  
/db\_xref="GI:21133410"  
/translation="WHLGRSRGRAVRSASVIMVLSGTIACPFVAQVATACDAQSVAA  
AAVAVAASGVGAAAVPARRAVAQCTCKPMPYDPSIRLAIAPAGCATTPAGERDIE

DEFINITION	Strepptomyces hygroscopicus AHBA gene cluster, partial sequence.
AF506520	
AF506520.2	GI:39932993
ACCESSION	Strepptomyces hygroscopicus
VERSION	Strepptomyces hygroscopicus
KEYWORDS	Bacteria; Actinobacteria; Actinomycetales;
SOURCE	Streptomycinae; Streptomycetaceae; Streptomycetes.
ORGANISM	
REFERENCE	
AUTHORS	1 Gao,Q. and Wang,Y.
TITLE	Identification and analysis of two separate AHBA biosynthetic gene clusters in Streptomyces hygroscopicus 17997
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 13513)
AUTHORS	Wang,Y. and Gao,Q.
TITLE	Direct Submission
JOURNAL	Submitted (26-APR-2002) Microbial Pathway Engineering, Medicinal Biotechnology, Tiantan, Beijing 100050, China
REFERENCE	3 (bases 1 to 13513)
AUTHORS	Gao,Q. and Wang,Y.
TITLE	Direct Submission
JOURNAL	Submitted (31-MAR-2003) Microbial Pathway Engineering, Medicinal Biotechnology, Tiantan, Beijing 100050, China
REMARK	Sequence update by submitter
COMMENT	On Dec 16, 2003 this sequence version replaced gi:24637548.
FEATURES	Location/Qualifiers
source	1..13513
	/organism="Streptomyces hygroscopicus"
	/mol_type="genomic DNA"
	/strain="17997"
	/db_xref="taxon:1912"
	1..699
	/note="phosphatase"
	/codon_start=1
	/transl_table=1
	/product="Shp"
	/protein_id="AAG3810.1"
	/db_xref="GI:24637549"
	/translation="MTRGPEFAAPTEIQRPPVRAHVFIDLDGVNDSPEWGEAFS
	LAVAEVGTGEAPPEEYRHRQRYEPDIMGLEPMEEPPVRESYLRDVGVYQGG
	VLYVRLTNRNGRLAATATGKAGRRASLLDVGILPYPAHVIGSDVPRPKPAPDII
	RRALBLEVPFERAIIIMTGDAPTDLASAHGADVTRAAALMCGQSEALLAADPDVYLRK
	PADLLALCPALPGH"
	638..1714
	/note="ShnK"
	/codon_start=1
	/transl_table=1
	/product="AHBA kinase"
	/protein_id="AAR32671.1"
	/db_xref="GI:39932994"
	/translation="MSSCGGPPPTCSPEARPCOATPEPPVAAAPSPMHLGIDIGRKVAL
	RASGADRTREHTLMPNDDODGPDRLALISDVAALAAAPTGQFOKAVGAVATVTPRP
	GERVTSPPSPSTGLALGPALRALVYGAALINAMDGCVALLAERAAADANLLYVY
	GTYGGGLVIGRLCPGPGRGSPFLGALLVALGQVCGRQGLQAIASGAPATLARA
	RELHGEVTEYELREALRTARPMATTAIDTCALAAIANGVELLHPGVIAGGFTPT
	AGIDNIVDRIATLHLPMSRRGPRAPAPAPARLGLSRLGAVILARAVTGAVSPVAA
	AAPRGDRGAASSTPPALHQEA"
	1715..3202
	/note="Shn01"
	/codon_start=1
	/transl_table=1
	/product="cyclohex-1-ene-1-carboxylate CoA ligase"
	/protein_id="AAR32672.1"
	/db_xref="GI:39932995"
	/translation="MSWSRLADYDEQFASALRVILVGPGQVAVVQLPMNMTSALLIA
	VMRGAVAVPITALRSELEKCLAQRLAVVMVTGVGWYDAGALAAEAERLPHLR
	HRVAGDAGGEMDFQDHFQPTPMQAQYMPDPEPDQVAVLFTAGTGGKGV
	LHSTNSDASAAGAIVGEERISGEASATPSLISLISDVLSPVYSGRAVLFDKM
	APRGIAIVAEARVSIITTAASPTDMSVRAVAANGDSTSLRYLFSGCAIIPKVAE
	VYRELIALAATYAGMTLSIGITRTSPDPDMAHSDGRPAALIDLRADGYSKAN
	PAVILYRAGACVAVGLDTGAVEIILCDRDGMLDTGDLVDPGRGGIRIVGRVADI
	TAQGMLEPVNDVEAGVLEHPAVAAVAIVATGDAVGETGCAVIVAAQPPPTLEQRLTYL
	TGQMEAYOPTLEVEVPLPRNANGVLRREILRORVOESGHBA"



REFERENCE	Pseudomonadaceae; Pseudomonas.
AUTHORS	1 (bases 1 to 15356) Stover, C.K., Pham, X.-O.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Hutnagle, W.O., Kowalik, D.U., Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Lardi, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z. and Paulsen, I.T.
TITLE	Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
JOURNAL	Nature 406 (6799), 959-964 (2000)
MEDLINE	20437337
PUBMED	10984043
REFERENCE	2 (bases 1 to 15356) Stover, C.K., Pham, X.-O.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Hutnagle, W.O., Kowalik, D.U., Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Lardi, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sater, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.
TITLE	Direct Submission
JOURNAL	Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 351145, Seattle, WA 98195, USA
REFERENCE	3 (bases 1 to 15356) Pseudomonas aeruginosa Community Annotation Project (PseudocAP) Direct Submission Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada
AUTHORS	This represents the February 3, 2003 version of the continually updated, reviewed, Pseudomonas aeruginosa PAO1 genome annotation, from PseudocAP (see <a href="http://www.pseudomonas.com">http://www.pseudomonas.com</a> for latest updates and links to alternate annotations). PseudocAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submissions through <a href="http://www.pseudomonas.com">www.pseudomonas.com</a> of any proposed changes.
COMMENT	----- 'Protein name confidence' is used to rate our confidence of the accuracy of the protein name. Class 1: Function experimentally demonstrated in P. aeruginosa. Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known). Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene. Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences. -----
FEATURES	-----
SOURCE	Location/Qualifiers 1..15356 /organism="Pseudomonas aeruginosa PAO1" /mol_type="Genomic DNA" /strain="PAO1" /db_xref="taxon:208964" 83..1207 /gene="pilC" /locus_tag="PA4527" 83..1207 /gene="pilC" /locus_tag="PA4527" 83..1207 /gene="pilC" /locus_tag="PA4527" /note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane protein; Subcellular localization confidence: Class 1; Alt protein name: pilin biogenesis protein PilC" /codon_start=1 /transl_table=1 /product="eF11 frameshift type 4 fimbrial biogenesis
gene	
CDS	

```

gene
protein P11C"
/locus_tag="AAG07915.1"
/db_xref="GI:9950770"
/translation="MLVKAHRLKQGIPLAVYRKKGISLDAKKGVKVRPMIDALFPRQMA
TMKGVGVGLDSFDIGEGPDNPNNKRLVDEIKQEVSSGNSLANSLRKKQYIPDELTC
NLVAVGEGSGAGLEMLNDLRVATYKRTESLAKLRKANTYPIAVIVALLIVSALLIKR
NVFVAGSOGGAEPLAFQTMVNVNLSEFLQEMVAVIVGVALIGFTKEHLAKRLLKVR
PDLTKDITLPIFGIYKKSAAVARYATLSTTFAAGVPLDALDSVGAGNIVFKNA
VSKIKQDSTCOMLNFMSRTSVFPPNMAIQMTALIGSSGSLDENLSVASYEEVEVDN
AVNDITLMEPMINAVLGVNLVAMYLPIQLGNVVC"
1211..2083
/gene="p11D"
/locus_tag="PA4528"
/note="synonym: xcpa"
1211..2083
/gene="p11D"
/locus_tag="PA4528"
/note="Protein name confidence: Class 1 (function
experimentally demonstrated in P. aeruginosa); Subcellular
localization: Inner membrane protein; Subcellular
localization confidence: Class 1, Alt protein name: type 4
fimbrial biogenesis protein p11D"
/codon_start=1
/transl_table=1
/product="Type 4 prepilin peptidase p11D"
/protein_id="AAG07916.1"
/db_xref="GI:9950771"
/translation="WEPLIDYLAHPALFVLTCTILGLLVGSPFLNVVYHRLPKMERKRW
KAAEALAGLEPERKQATVNLVLPNSACPFGCHIRPMENPLVSYATLGGKSSCKA
ALIGKYPVEIATLALSGYVAMHGFPTQAGAMLLTWMGLAMSLIDADHQLPDLVLA
PIKRGYLIANHPGLFASLDALFGAVGYSTLSVFMFLVLYGKEGAGYDPEKLA
MIGANGGMOILLPTILLSLVGALLGVIMLRNRAESGTPIPPEPIYALAGMALANG
DOITRTYLOFPGFK"
2080..2691
/gene="coae"
/locus_tag="PA4529"
2080..2691
/gene="coae"
/locus_tag="PA4529"
/note="Protein name confidence: Class 2 (high similarity
to functionally studied protein); Alt protein name: Coae
protein"
/codon_start=1
/transl_table=1
/product="dephosphocoenzyme A kinase"
/protein_id="AAG07917.1"
/db_xref="GI:9950772"
/translation="MTQPMILGLTGIGSGKSAABEPIISGLVLDADNAAKRVNPE
GRPALAKYVERGCGIILPDQDLRALRRIIPAPREERWLBOLPLIGATIVOLYI
AASAEPAVLIVSLVESGQRQMTHRVAVVDPTEPHLOLQRTMLDKSSECVRSILQD
QARDERKRLHADVLVNDGDLHLQREVERLHAYTLRGGRA"
2688..2888
/locus_tag="PA4530"
2688..2888
/locus_tag="PA4530"
/note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG07918.1"
/db_xref="GI:9950773"
/translation="MSQPLTVECPGAPVEMKSDNRYRPFCSDRCLIDIGAAWAE
HAIPDGLDDIIFSDLPPEHH"
complement(2925..3134)
/locus_tag="PA4531"
complement(2925..3134)
/locus_tag="PA4531"
/note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1

```

/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07919.1"  
/db\_xref="GI:9950774"  
/translation="MGVHMAEGLRELLSKSDEADRAMRLALLSAIVEINRAERDA  
GDIASELOFLAENLDDQRDYAFMRP"  
complement(3240..3929)  
/locus\_tag="PA4532"  
complement(3240..3929)  
/locus\_tag="PA4532"  
/note="Protein name confidence: Class 4 (homologs of  
previously reported genes of unknown function, or no  
similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07920.1"  
/db\_xref="GI:9950775"  
/translation="WTAABLSAESLEEGMALIYAVVLLAAMRAFWELVSDTRQHL  
VFGTGLGFLMLVRRDESGLSFHFIGMTAVTLLDPLAIWGLVAQLGCLLGRQ  
DMLAIGVGVLLVLAIPALIAELALFVKRQPRNLFVIFCCGFPALITAVVLLG  
LGVWIDGLFPMPLIEDPAGYLMIMPEAFINGTVITGVVYVPEWLETFNRIRYL  
QAPKDEERKP"  
complement(3926..4396)  
/locus\_tag="PA4533"  
complement(3926..4396)  
/locus\_tag="PA4533"  
/note="Protein name confidence: Class 4 (homologs of  
previously reported genes of unknown function, or no  
similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07921.1"  
/db\_xref="GI:9950776"  
/translation="MKRLHWVLLAMLLGLAIVKSGSTLEPWTLODPEKASINDS  
TRVLVARDMSGKLVKALAEPRGYBARALTLADISMPALISRLFVAPNARDY  
SYPLIDRDAAIASRYASDEKVLWVRLQRRVLTQLDSVALIALDQLP"  
complement(4393..4818)  
/locus\_tag="PA4534"  
complement(4393..4818)  
/locus\_tag="PA4534"  
/note="Protein name confidence: Class 4 (homologs of  
previously reported genes of unknown function, or no  
similarity to any previously reported sequences)"

Query Match 78.3%; Score 18.8; DB 1; Length 15356;  
Best Local Similarity 70.8%; Pred. No. 4.8e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGCGCGSAGCSGAC 24  
|||||:||||:||||:||||:  
Db 11885 GTCTGTGCGCGCGCACCCGCAC 11862

RESULT 23  
AY258138 34398 bp DNA linear BCT 02-JAN-2004  
LOCUS Pseudomonas aeruginosa strain C gene island PA01-4(C) sequence.  
DEFINITION  
ACCESSION AY258138  
VERSION AY258138.1 GI:33114121  
KEYWORDS  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
Pseudomonas aeruginosa  
1 (bases 1 to 34398)  
AUTHORS Klockgether, J., Reva, O.N., Larbig, K. and Tummler, B.  
TITLE Sequence Analysis of the Mobile Genome Island pKlC102 of  
Pseudomonas aeruginosa C  
JOURNAL J. Bacteriol. 186 (2), 518-534 (2004)  
PUBMED 14702321  
REFERENCE 2 (bases 1 to 34398)

AUTHORS Klockgether, J., Reva, O.N., Larbig, K.D. and Tummler, B.  
TITLE Direct Submission  
JOURNAL Submitted (19-MAR-2003) Klinische Forschungsgruppe, Medizinische  
Hochschule Hannover, Carl-Neuberg-Strasse 1, Hannover 30623,  
Germany

FEATURES  
source  
1..34398  
location/Qualifiers  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="C"  
/isolation\_source="airways of a cystic fibrosis patient"  
/db\_xref="taxon:287"  
/map="between genes PA0976 and PA1003"  
/clone="C"  
/note="hypervariable genome region"  
1..298  
/note="similar to Pseudomonas aeruginosa PA01 genomic  
region containing gene PA0976"  
223..298  
/product="tRNA-Lys"  
299..23693  
/note="gene island PA01-4(C)"  
complement(423..551)  
/gene="CL-1"  
complement(423..551)  
/gene="CL-1"  
/note="similar to PA0977; possibly truncated"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAP94679.1"  
/db\_xref="GI:33114122"  
/translation="MLAEPRSFKNTAVPSVIADYVEGALLRLQNGFRALETEKPT"  
complement(669..1949)  
/gene="xerc"  
/note="CL-2a"  
complement(669..1949)  
/gene="xerc"  
/note="phage-like integrase; similar to xerc-integrase  
from Pseudomonas aeruginosa plasmid pKlC102"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="putative integrase"  
/protein\_id="AAP94680.1"  
/db\_xref="GI:33114123"  
/translation="MTPQQLTEVYIPAPDLREASAKTYRATKALIKHPPTATVOEV  
DHRVTLGRRKRYLEQGLSKSMWTYSNLRITNGYATIEHLYTHSYNPNPKTIVIP  
RRASKTYAAEALILRRWMLNQVGAERCTGDRATTPAWMLCTFEVYVYIGIRLNL  
LCIRKRDIDWNQILIRGETEKHKERVYITGLVPHLSRLQLQEDMDRAFDADDL  
FNVRSPHYKSNSDOVAMRYKLEKGVMTPIRFRHTLATDLMKAPBNIIHL  
TKCLNSNIOTTSYIEADYDHRVAVHARSLAQGLAENYKVDYSGSPASAKPKP  
CGQPLARMGVYPPPEAREPTEPREHITPGTIOGGPYREBALQPPPTPQSVLFTL  
MAQHLNRAASASAAAPATSGSGGWSGTARSSLA"  
1353..2042  
/gene="CL-2b"  
1353..2042  
/gene="CL-2b"  
/note="innerorf; similar to Pseudomonas aeruginosa  
plasmid pKlC102 ORF CP103b"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="putative excision regulator"  
/protein\_id="AAP94702.1"  
/db\_xref="GI:33114145"  
/translation="WRHNPRLRYWYELFMSFLCIADEDQIIPVDVALADAQGRVOP  
DAGVVEDIKRAKPEPGSDACAIGALAGADLHYEPILAQADGCGDGFACSPGDDDG  
GLSRVVDKRAVGHOLVLDGIIPDRSOMRIRVPVAPROALFODLAPPSQHRPVVHLVG  
CRRAVLEQRIGRGAVDRAAGFSRLVYEDVLLGELLRHHSITPPASAVTVALGSRGC  
TGNSSLSGWSR"  
complement(1946..3772)

gene  
CDS



CDS  
 /gene="CL-3"  
 complement(1946. .3772)  
 /note="similar to ORF CP102 from *Pseudomonas aeruginosa*  
 plasmid pKLC102"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAP94681.1"  
 /db\_xref="GI:33114124"  
 /translation="MELIGPRRRLLENTIMQASLSKQPEETIYRRPIANVAYVOO  
 LPASENHHHAPGMDHGLEIYVAYALKVRTYLLPIGAAESSQAQAEASAAAG  
 ALANDIGTIVDLEIJDGSTMHPNGEPIQPYRFKTVSRBYQLGASALIHOL  
 LPTALDWLSRFPBLMAQLYLFAQYEHAIIGELIYKADQASVAQELGSPRALA  
 AKPSQLORLADGLFLVKDFKLNOPGSPDGWLTODALMLVSKPADQIRAVILAQ  
 GIEGVSSNAPFFNMLQDAVLOTNAEDKAIWTATINDGAKRNFLLKAPLIVA  
 DPARRPDSYSGSLVVEGNSPEKPEITCEIPNDPIORORAPAKMTLRDAPITLAP  
 SNDIQAIKAPSTDQERTDDIYALGINSPPELSTDSHPANVPANTRRENTQOP  
 LGKEPIDCAPEALIDVFMPSRTDLGGFVDMKSGITARLFINDKALVHTVDGT  
 AMLVTPGIFKRYVQEHPEVEKLAQAKTAGKLVORAFKQGLHRTKTSKNLTIWITIV  
 SGPRKTKELKAYLLQDPKTLFPVQPLNDPSLTVITDAEGVE"  
 complement(4196. .4546)  
 /gene="CL-4"  
 complement(4196. .4546)  
 /note="similar to plasmid stabilization system protein  
 ParB; similar to ORF CP93 from *Pseudomonas aeruginosa*  
 plasmid pKLC102"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAP94682.1"  
 /db\_xref="GI:33114125"  
 /translation="MAKTRISHDAQADIDIRFTRNFGDAARRVQALIGALEAV  
 ASDQVGSISREELGAGLRSHILVYCKSMNVKAVRPRHFVFRVATQVLEVVY  
 LHMDSMDLOHLPHR"  
 complement(4550. .4789)  
 /gene="CL-5"  
 complement(4550. .4789)  
 /note="similar to predicted transcriptional regulators  
 containing the CopG/ArcMecJ DNA-binding domain (COG  
 3609); similar to ORF CP92 from *Pseudomonas aeruginosa*  
 plasmid pKLC102"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAP94683.1"  
 /db\_xref="GI:33114126"  
 /translation="MEODINELVETGYQRNSEYIRAGRLLLQOEONSAKLEALN  
 ATSSGLAQLEGERDELSDLAQYDLBELGQASH"  
 complement(5311. .6822)  
 /gene="CL-6"  
 complement(5311. .6822)  
 /note="similar to ORF CP91 from *Pseudomonas aeruginosa*  
 plasmid pKLC102"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAP94684.1"  
 /db\_xref="GI:33114127"  
 /translation="MTMTNDYLYEYVLTLLGMIINNGIMNMISDTGLFAVPFAVYVR  
 EMLKVRGADBNKGVLSARIEHTIYVIGYIVVALGIPVNVSPDITEPDARAOO  
 COVYLPAADTGSTFSLSLAKSCQAPIMAMMHASIKETGAAVAATACGDTIRM  
 RMEYDNRVNNPLLAQEIADPSRCYCPSPRRLEFMRPPDGLSVCEDKALADLNTIS  
 RFLNLTGAYDIDYSKPSROSWPYNATRDGLPVGAGGAGPTCKQWMAOSGILROR  
 IKQGVDDMLNTSFLKMAKMDQDEVTAVIRQVISPSSQYKGVNTYDYGQGVGTWVA  
 GIATAGTGFVALGSLAVFPAMDVVRQALPMVMSFLKMAVITCIPMVLVIGTYQLKVA

CDS  
 /gene="CL-7"  
 complement(6819. .7166)  
 /note="similar to ORF CP90 from *Pseudomonas aeruginosa*  
 plasmid pKLC102"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAP94685.1"  
 /db\_xref="GI:33114128"  
 /translation="MADTLRSRLKQLQGLVGVNIVVGLAVGTLLSLPALNRRGIQO  
 LEAMQSNVMSLPAWRALYLCALAIAMFRILKQKKEISAHRRQIRRIEILVLLILLI  
 EFSKAYRTGGAA"  
 complement(7166. .8548)  
 /gene="CL-8"  
 complement(7166. .8548)  
 /note="similar to ORF CP89 from *Pseudomonas aeruginosa*  
 plasmid pKLC102"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAP94686.1"  
 /db\_xref="GI:33114129"  
 /translation="MNITTSVALTWLAAQLAQDDPINNSKTGTVLSDEVLYSIGGS  
 AVMSGSAQMDISIGVGWNNNDLGMNISTYLENQNLGATGSPONIKESVIONATG  
 AVMSLPALIORANPOLYNLTNGILQARDYDRSGKTCAIAEKADIEBQTMCK  
 IAGQALGATLASGDQAVSALFVEKKGNDQVTWGGDKAGSGGQPIRIYNDVTR  
 IAGNLTLSRVDSSSVSPSACNNGVCMWSPQBAAPATVGLGEOOQOCEGCK  
 TWTAAAGTLPILQETYDKLQSLQELSKRELTENLAACTDILLPIRGVTELR  
 DERDQVTLARILSDVSLNDVSKALLLRLNFRAGAKEPVANAAGLATQAVDQTSIL  
 QOEISNLKTELEIRRELIASNPVRVIERGQORASGSSGVESA PADRDRLQAPSA  
 GKSGSGCRP"

Query Match 78.3%; Score 18.8; DB 1; Length 34398;  
 Best Local Similarity 70.8%; Pred. No. 4.3e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 GTCCGTGTGCGCGTSACSGSAC 24  
 Db 18949 GTCCGCTGCGCGTCACCCCAC 18972

RESULT 24  
 AF271693/c 50550 bp DNA linear PHG 28-JAN-2001  
 LOCUS AF271693  
 DEFINITION Mycobacteriophage Bxb1, complete sequence.  
 ACCESSION AF271693  
 VERSION AF271693.1 GI:12583995  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mycobacteriophage Bxb1  
 Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
 REFERENCE  
 1 (bases 1 to 50550)  
 Mediavilla, J., Jain, S., Kriakov, J., Ford, M.E., Duda, R.L.,  
 Jacobs, W.R., Jr., Hendrix, R.W. and Hatfull, G.F.  
 Genome organization and characterization of mycobacteriophage Bxb1  
 Mol. Microbiol. 38 (5), 955-970 (2000)  
 MEDLINE 20572070  
 PUBMED 11123671  
 REFERENCE 2 (bases 1 to 50550)  
 Mediavilla, J., Jain, S., Kriakov, J., Ford, M.E., Duda, R.L.,  
 Jacobs, W.R., Hendrix, R.W. and Hatfull, G.F.  
 Direct Submission  
 Submitted (24-MAY-2000) Biological Sciences, University of  
 Pittsburgh, Ruben Hall, Pittsburgh, PA 15260, USA  
 FEATURES  
 source 1. .50550

gene /organism="Mycobacteriophage Bxb1"  
CDS /mol\_type="genomic DNA"  
599..895 /db\_xref="taxon:148603"  
/gene="1"  
599..895  
/note="related to L5 gp4 (47%); 11.4 kD hypothetical  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp1"  
/protein\_id="AAG59706.1"  
/db\_xref="GI:12583996"  
/translation="MGWESSDRGRRLPADWTRIRREVLRAADHRCQIRYADICIMAT  
EVDHYRVEDDEEPLQASCKPCARKSAMEGVAQRAKLRAKKRPPPRHGRSS"  
930..1370  
/gene="2"  
930..1370  
/note="related to L5 gp5 (41%); 16.3 kD hypothetical  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp2"  
/protein\_id="AAG59707.1"  
/db\_xref="GI:12583997"  
/translation="MGTGRLPNRDERVRNKEEYGVTTLLPVSGPVKSPPLGLTDP  
HPIVDLYNSLAESNAALYOPSDMHWAKFTLLHFPADOLLKSSKSSQLVAVNQMLSS  
LIVSGDRRRVTEVERTKSDGPDASVTTMGBLFRALKRKSS"  
1716..2027  
/gene="3"  
1716..2027  
/note="11.4 kD hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp3"  
/protein\_id="AAG59708.1"  
/db\_xref="GI:12583998"  
/translation="MAVVTYDNGQEWVHGHKXETDPDPOSILLIISNOGEVYGF  
NAQSWTHFRVPMQIVRIGDITLRKDGKMAVIVGDSYEPFGHITFPAEEVAEKGLV  
V"  
2291..3361  
/gene="4"  
2291..3361  
/note="related to L5 gp31 (43%); 60aa; 36.7 kD  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp4"  
/protein\_id="AAG59709.1"  
/db\_xref="GI:12583999"  
/translation="MTVTITADVRLDTGQPDNQGVFSTVLRLQDGSILLITQKQVNP  
VGAALSVLEPGFAIVVGEYFIEVEPTDAGLWGLATSAVAVPDTSAELAAVAVN  
GIDANPPGADMDALSNVSEPPPSADHDVLDLSAIAAYLASNPBAGSVAMD  
DIDKSTFTPSSTHSIANVYGLQDALDEKIDRAVAVARSLGTAALVDASPTLDT  
LAEIALAALDDEPRFATTVASQIGARADKATTTTAGTGLTGGCDLSANRTIYNSRSTSS  
TTACVGNDSRLSNTRIPTDGSVTNAKVASGAGIALSKLATGTVAGSDNSGANTLTITW  
GTEAQTATGTDNSNTIYLRTA"  
3371..4081  
/gene="5"  
3371..4081  
/note="related to L5 gp32 (31%); 60aa; 23.3 kD  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp5"  
/protein\_id="AAG59710.1"  
/db\_xref="GI:12584000"  
/translation="MAGMSLATTAFAKAAIGSTLEIQKISIGTTEIWSAAPPTVDFDA

gene VVSFNSALGDLVSFSATAGARVFAVHLGNNTVASVTYGGNAMLVTGAECTMNS  
CDS ASSGIVLAVYTLASPGSGQTVENDRNANMAASPAISYLVASVGTASTAGCTSS  
HSISAPPANGRVFQAFDMSSSNVSLTTPSGGTGLNRDGTAGMAARDSAAISTGTI  
SSTGAPVASIAIPMTPTV"  
4095..4355  
/gene="6"  
4095..4355  
/note="9.7 kD hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp6"  
/protein\_id="AAG59711.1"  
/db\_xref="GI:12584001"  
/translation="MTVTATVADHIDSGRPDDSHHTSSDLREQGVITTRVAVKPF  
NGEIALTLPPQPAVTVTHRODRLIDVEEDSDMLDLLEAATD"  
4363..4620  
/gene="7"  
4363..4620  
/note="related to TM4 gp90 (23%); 9.7 kD hypothetical  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp7"  
/protein\_id="AAG59712.1"  
/db\_xref="GI:12584002"  
/translation="TNNRLITMFAALYKAVEDYLRAHPEFLNQVIDRATAMPDLADL  
DDKLIATKIPDLRDDKTIIGWFPDLRLPEQLINAINPFKA"  
4620..6053  
/gene="8"  
4620..6053  
/note="related to D29 gp10; H11415; 53.3 kD hypothetical  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp8"  
/protein\_id="AAG59713.1"  
/db\_xref="GI:12584003"  
/translation="MPRVVYGLTSSNGMPYLNDSGCEWTKIPGTSVTLQIAKQPLA  
ILRAPADPHAVYERLDADACMTPTNSVPSNHLGSLAMDLMNTHPEQVPAGFD  
AAKARVKEILDITFEGKVPFNGNDSSPRDANHPLQASLRNGTDEDTGNNPTADPIAR  
KIRADGSTERFGSAPASAPITLAAATGTSARAAETIDPAVSGLRBSECTNVRIAM  
WIAQIGHSGSFQYTEBIANKGRVAPYIGRTWQITMDYNRSFSQWAVAGVPTPD  
YFVNVYREIADLKWAGIGPAMVTVAPDIDELSDRLNVTTRIRNGTGLADROA  
RYNRALQGDALLIHLHEEDPFLSALTDAEORELIDLARQOAKYKRSRSPILHWPBEG  
EVDITAGLSWSTDNVHIQIVYEKLAVITGVSVIALLYAVANSDDPTNNPELAKRIK  
RVKEDITAAQVOIQKMLAEQKFAHA"  
6043..7011  
/gene="9"  
6043..7011  
/note="related to L5 gp12 (38%); 35.9 kD hypothetical  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp9"  
/protein\_id="AAG59714.1"  
/db\_xref="GI:12584004"  
/translation="WPLKLGDRNPTVVRKREVMARFAGYARVHGPIPTDDEGPRRA  
EAMQVEYSRTFQPLDGIIVSDDLRSLSGI PAPEADRPVLLVSGTGVPMVGPADVA  
RRLGDIYLMRPVGPYTAQAFPMGDSVANGVETATRLIEEBERRIERGLSMIGSOG  
AIVTSELMEYHIKPVYGRLHWKDRGAATFGNPMRETGKVPDPGCGMSAKSHGI  
ADQLMVDTPDMWNRVYAHKGDLYTDCGSGSGMKTAIYKVMMASRVFSGPDSILRLQLE  
IGVNPTEFLIALIRAVLDAGLFTIRGTGPHVNNIDPATDILRSVT"  
7031..8719  
/gene="10"  
7031..8719  
/note="related to L5 gp13 (65%); 62.3 kD putative  
terminase protein"  
/codon\_start=1



```
* 1158 1257: gap of unknown length
* 1258 2321: contig of 1064 bp in length
* 2322 2421: gap of unknown length
* 2422 3618: contig of 1197 bp in length
* 3619 3718: gap of unknown length
* 3719 4930: contig of 1212 bp in length
* 4931 5030: gap of unknown length
* 5031 6031: contig of 1001 bp in length
* 6032 6131: gap of unknown length
* 6132 7270: contig of 1139 bp in length
* 7271 7370: gap of unknown length
* 7371 8508: contig of 1138 bp in length
* 8509 8608: gap of unknown length
* 8609 10588: contig of 1980 bp in length
* 10589 10688: gap of unknown length
* 10689 12189: contig of 1501 bp in length
* 12190 12289: gap of unknown length
* 12290 14285: contig of 1996 bp in length
* 14286 14385: gap of unknown length
* 14386 15907: contig of 1522 bp in length
* 15908 16007: gap of unknown length
* 16008 17891: contig of 1884 bp in length
* 17892 17991: gap of unknown length
* 17992 19622: contig of 1631 bp in length
* 19623 19722: gap of unknown length
* 19723 21796: contig of 2074 bp in length
* 21797 21896: gap of unknown length
* 21897 24606: contig of 2710 bp in length
* 24607 24706: gap of unknown length
* 24707 26618: contig of 1912 bp in length
* 26619 26718: gap of unknown length
* 26719 28720: contig of 2002 bp in length
* 28721 28820: gap of unknown length
* 28821 31309: contig of 2489 bp in length
* 31310 31409: gap of unknown length
* 31410 34084: contig of 2675 bp in length
* 34085 34184: gap of unknown length
* 34185 36260: contig of 2076 bp in length
* 36261 36360: gap of unknown length
* 36361 37424: contig of 1064 bp in length
* 37425 37524: gap of unknown length
* 37525 39771: contig of 2247 bp in length
* 39772 39871: gap of unknown length
* 39872 43194: contig of 3323 bp in length
* 43195 43294: gap of unknown length
* 43295 46776: contig of 3482 bp in length
* 46777 46876: gap of unknown length
* 46877 51203: contig of 4327 bp in length
* 51204 51303: gap of unknown length
* 51304 53821: contig of 2518 bp in length
* 53822 53921: gap of unknown length
* 53922 56371: contig of 2450 bp in length
* 56372 56471: gap of unknown length
* 56472 60412: contig of 3941 bp in length
* 60413 60512: gap of unknown length
* 60513 65444: contig of 4932 bp in length
* 65445 65544: gap of unknown length
* 65545 71808: contig of 6264 bp in length
* 71809 71908: gap of unknown length
* 71909 81306: contig of 9398 bp in length
* 81307 81406: gap of unknown length
* 81407 94058: contig of 12652 bp in length.
```

## FEATURES

```
source 1..94058
        /organism="Rattus norvegicus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10116"
        /clone="CH230-482H23"
```

## ORIGIN

```
Query Match      78.3%;   Score 18.8;   DB 2;   Length 94058;
Best Local Similarity 70.8%;   Pred. No. 3.7e+03;
Matches 17;   Conservative 5;   Mismatches 2;   Indels 0;   Gaps 0;
```

```
QY      1  GTCTGSGTGGCCGTGACGCGSAC 24
Db      9282  GTCTGGGGGCGCGGAGCGCGAC 9259
```

RESULT 26  
AP006618\_13  
WPCOMMENT

Sequence split into 61 fragments LOCUS AP006618 Accession AP006618

Fragment Name	Begin	End
AP006618_00	1	110000
AP006618_01	100001	210000
AP006618_02	200001	310000
AP006618_03	300001	410000
AP006618_04	400001	510000
AP006618_05	500001	610000
AP006618_06	600001	710000
AP006618_07	700001	810000
AP006618_08	800001	910000
AP006618_09	900001	1010000
AP006618_10	1000001	1110000
AP006618_11	1100001	1210000
AP006618_12	1200001	1310000
AP006618_13	1300001	1410000
AP006618_14	1400001	1510000
AP006618_15	1500001	1610000
AP006618_16	1600001	1710000
AP006618_17	1700001	1810000
AP006618_18	1800001	1910000
AP006618_19	1900001	2010000
AP006618_20	2000001	2110000
AP006618_21	2100001	2210000
AP006618_22	2200001	2310000
AP006618_23	2300001	2410000
AP006618_24	2400001	2510000
AP006618_25	2500001	2610000
AP006618_26	2600001	2710000
AP006618_27	2700001	2810000
AP006618_28	2800001	2910000
AP006618_29	2900001	3010000
AP006618_30	3000001	3110000
AP006618_31	3100001	3210000
AP006618_32	3200001	3310000
AP006618_33	3300001	3410000
AP006618_34	3400001	3510000
AP006618_35	3500001	3610000
AP006618_36	3600001	3710000
AP006618_37	3700001	3810000
AP006618_38	3800001	3910000
AP006618_39	3900001	4010000
AP006618_40	4000001	4110000
AP006618_41	4100001	4210000
AP006618_42	4200001	4310000
AP006618_43	4300001	4410000
AP006618_44	4400001	4510000
AP006618_45	4500001	4610000
AP006618_46	4600001	4710000
AP006618_47	4700001	4810000
AP006618_48	4800001	4910000
AP006618_49	4900001	5010000
AP006618_50	5000001	5110000
AP006618_51	5100001	5210000
AP006618_52	5200001	5310000
AP006618_53	5300001	5410000
AP006618_54	5400001	5510000
AP006618_55	5500001	5610000
AP006618_56	5600001	5710000
AP006618_57	5700001	5810000
AP006618_58	5800001	5910000
AP006618_59	5900001	6010000
AP006618_60	6000001	6021225

Continuation (14 of 61) of AP006618 from base 1300001 (AP006618 Nocardia farcinica ITM 1

Query Match 78.3%; Score 18.8; DB 1; Length 110000;  
 Best Local Similarity 70.8%; Pred. No. 3.6e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCCGATGCGCCGATCGGAC 24  
 Db 86535 GTCCGATGCGCCGATCGGAC 86535

RESULT 27  
 LOCUS CNS08CAM 146585 bp DNA linear PLN 21-NOV-2003  
 DEFINITION Oryza sativa chromosome 12. BAC OJ1268\_D02 of library Monsanto  
 from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza  
 sativa (rice), complete sequence.  
 AL831809  
 ACCESSION AL831809.3 GI:25900557  
 VERSION HTG  
 KEYWORDS Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 146585)  
 AUTHORS Cholsine, N., Orjeda, G., Catolico, L., Demange, N., Wincker, P.,  
 Segurens, B., Pelletier, E., Scarpelli, C., Salanoubat, M.,  
 Weissenbach, J., and Quetier, F.  
 Oryza sativa chromosome 12 sequencing  
 Unpublished  
 2 (bases 1 to 146585)  
 TITLE Genoscope.  
 REFERENCES Direct Submission  
 JOURNAL Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage :  
 BP 191 91006 Evry cedex - FRANCE (E-mail : segref@genoscope.cns.fr  
 On Nov 28, 2002 this sequence version replaced gi:24817645.  
 Center: Genoscope / Centre National de Sequencage  
 COMMENT Center code: GS  
 Web site: http://www.genoscope.cns.fr/  
 Contact: Segref@genoscope.cns.fr

The following sequence is oriented from the T7 to the SP6 end. The  
 nucleotide sequence of this BAC clone was generated by combining  
 Monsanto and Genoscope sequencing data.  
 Upstream BAC (overlapping the T7 end) : OJ1559\_C07 (AC=AL731887)  
 Downstream BAC (overlapping the SP6 end) : OJ1119\_E02 (AC=AL731762)  
 ----- Finishing boundaries  
 FINISHED SEGMENT STARTS AT BASE 1  
 FINISHED SEGMENT ENDS AT BASE 146585  
 -----

FEATURES  
 source location/Qualifiers  
 1. 146585  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /sub\_species="japonica"  
 /db\_xref="taxon:39947"  
 /chromosome="12"  
 /clone="OJ1268\_D02"  
 /clone\_1db="Monsanto"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 8; Length 146585;  
 Best Local Similarity 70.8%; Pred. No. 3.5e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCCGATGCGCCGATCGGAC 24  
 Db 102468 GTCCGATGCGCCGATCGGAC 102491

RESULT 28

AC084284/c  
 LOCUS AC084284 152186 bp DNA linear HTG 11-MAR-2001  
 DEFINITION Homo sapiens chromosome RPL1-11 clone RPL1-276D3, WORKING DRAFT  
 SEQUENCE, 30 unordered pieces.  
 AC084284  
 ACCESSION AC084284.3 GI:13270752  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 152186)  
 AUTHORS Waterston, R.H.  
 TITLE The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 152186)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-OCT-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On Mar 10, 2001 this sequence version replaced gi:10945792.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Project Information -----  
 Center project name: H\_NH0276D03  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 100%  
 Sequencing method: plasmid; 0%  
 Chemistry: Dye-terminator Big Dye; 0% of reads  
 Assembly program: Phrap; version 0.99019  
 Consensus quality: 132621 bases at least Q40  
 Consensus quality: 139102 bases at least Q30  
 Consensus quality: 142309 bases at least Q20  
 Insert size: 158000; agarose-fp  
 Insert size: 149286; sum-of-contigs  
 Quality coverage: 3.13 in Q20 bases; agarose-fp  
 Quality coverage: 3.40 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 30 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 1511: contig of 1511 bp in length  
 1512  
 1611: gap of unknown length  
 1612  
 2942: contig of 1331 bp in length  
 2943  
 3042: gap of unknown length  
 3043  
 4265: contig of 1223 bp in length  
 4266  
 4365: gap of unknown length  
 4366  
 6324: contig of 1959 bp in length  
 6325  
 6424: gap of unknown length  
 6425  
 7995: contig of 1571 bp in length  
 7996  
 8095: gap of unknown length  
 8096  
 9575: contig of 1480 bp in length  
 9576  
 9675: gap of unknown length  
 9676  
 11960: contig of 2285 bp in length  
 11961  
 12060: gap of unknown length  
 12061  
 14652: contig of 2592 bp in length  
 14653  
 14752: gap of unknown length  
 14753  
 17032: contig of 2280 bp in length  
 17033  
 17132: gap of unknown length  
 17133  
 19636: contig of 2504 bp in length  
 19637  
 19736: gap of unknown length  
 19737  
 23580: contig of 3844 bp in length  
 23581  
 23680: gap of unknown length

```
* 23681 26611: contig of 2931 bp in length
* 26612 26711: gap of unknown length
* 26712 29432: contig of 2721 bp in length
* 29433 29532: gap of unknown length
* 29533 34556: contig of 5024 bp in length
* 34557 34656: gap of unknown length
* 34657 37667: contig of 2911 bp in length
* 37668 41750: contig of 4083 bp in length
* 41751 41850: gap of unknown length
* 41851 47006: contig of 5156 bp in length
* 47007 52437: contig of 5331 bp in length
* 52438 52537: gap of unknown length
* 52538 57297: contig of 4760 bp in length
* 57298 57398: gap of unknown length
* 57398 62529: contig of 5132 bp in length
* 62530 62629: gap of unknown length
* 62630 66478: contig of 3849 bp in length
* 66479 75759: contig of 9181 bp in length
* 75760 75859: gap of unknown length
* 75860 83433: contig of 7574 bp in length
* 83434 83533: gap of unknown length
* 83534 93431: contig of 9897 bp in length
* 93431 100839: contig of 7309 bp in length
* 100840 100939: gap of unknown length
* 100940 109890: contig of 8951 bp in length
* 109891 109990: gap of unknown length
* 109991 119771: contig of 9781 bp in length
* 119772 119871: gap of unknown length
* 119872 128562: contig of 8691 bp in length
* 128563 128662: gap of unknown length
* 128663 140541: contig of 11879 bp in length
* 140542 140642: gap of unknown length
* 140642 152186: contig of 11545 bp in length.
Location/Qualifiers
1. 152186
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="RP11-276D3"
/clone="RP11-276D3"
1. 1511
/note="assembly_name:Contig11"
misc_feature 1612..2942
/note="assembly_name:Contig12"
misc_feature 3043..4265
/note="assembly_name:Contig13"
misc_feature 4366..6324
/note="assembly_name:Contig14"
misc_feature 6425..7995
/note="assembly_name:Contig15"
misc_feature 8096..9575
/note="assembly_name:Contig16"
misc_feature 9676..11960
/note="assembly_name:Contig17"
misc_feature 12061..14652
/note="assembly_name:Contig18"
misc_feature 14753..17032
/note="assembly_name:Contig19"
misc_feature 17133..19636
/note="assembly_name:Contig20"
misc_feature 19737..23580
/note="assembly_name:Contig21"
misc_feature 23681..26611
/note="assembly_name:Contig22"
misc_feature 26712..29432
/note="assembly_name:Contig23"
misc_feature 29533..34556
/note="assembly_name:Contig24"
misc_feature 34657..37567
```

```
misc_feature /note="assembly_name:Contig25"
37668..41750
/note="assembly_name:Contig26"
41851..47006
/note="assembly_name:Contig27"
47107..52437
/note="assembly_name:Contig28"
52538..57297
/note="assembly_name:Contig29"
57398..62529
/note="assembly_name:Contig30"
62630..66478
/note="assembly_name:Contig31"
66579..75759
/note="assembly_name:Contig32"
75860..83433
/note="assembly_name:Contig33"
83534..93430
/note="assembly_name:Contig34"
93531..100839
/note="assembly_name:Contig35"
100940..109890
/note="assembly_name:Contig36"
109991..119771
/note="assembly_name:Contig37"
119872..128562
/note="assembly_name:Contig38"
128663..140541
/note="assembly_name:Contig39"
140642..152186
/note="assembly_name:Contig40"
```

## ORIGIN

```
Query Match 78.3%; Score 18.8; DB 2; Length 152186;
Best Local Similarity 70.8%; Pred. No. 3.4e+07;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 GTCCTGCTGCGCCGTACGCGCAC 24
Db 95430 GACCTGTGTGCGCCGTGCGCGCAC 95407
```

```
RESULT 29
AL663032 162995 bp DNA linear ROD 05-APR-2002
LOCUS Mouse DNA sequence from clone RP23-2716 on chromosome X, complete
DEFINITION
VERSION AL663032.9 GI:20068678
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Whitehead,S.
TITLE Direct SubMISSION
JOURNAL Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
hunquerry@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:19335849.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
```

assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Mp., MORMPEP; Information on the MORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-2716 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/dacpac/home.htm> VECTOR: pBAC3.6.

# FEATURES

## source

```
1.162995
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="X"
  /clone="RP23-2716"
  /clone_lib="RPI-23"
```

## ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 162995;  
Best Local Similarity 70.8%; Pred. No. 3.4e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```
Qy 1 GTCCTGTGGCCGCTGACGCGAC 24
    |||||:|||||:|||||:|||||:
Db 158767 GTCCGGGGGGCCGGGACGCGCAC 158790
```

```
RESULT 30
AL646062 193050 bp DNA linear BCT 02-SEP-2002
```

DEFINITION Ralstonia solanacearum GMI1000 chromosome, complete sequence;  
segment 6/19

```
ACCESSION AL646062
VERSION AL646062.1
KEYWORDS GI:17427974
```

## SOURCE

ORGANISM Ralstonia solanacearum  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Burkholderiaceae; Ralstonia.

## REFERENCE

1 Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,  
Ariat,M., Billault,A., Broctier,P., Camus,J.C., Catolico,L.,  
Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,  
Gaepin,C., Lavie,M., Moisan,A., Robert,C., Saurin,M., Schiek,T.,  
Sigulier,P., Thebaul,P., Whalen,M., Wincker,P., Levy,M.,  
Weissenbach,J. and Boucher,C.A.  
Genome sequence of the plant pathogen Ralstonia solanacearum  
Nature 415 (6871), 497-502 (2002)

TITLE JOURNAL  
MEDLINE 21681879  
PUBMED 11823852  
2 (bases 1 to 193050)  
Boucher,C.A.  
Direct Submission  
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMM CNRS  
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire  
INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
<http://sequence.toulouse.inra.fr/R.solanacearum.html>.

## REFERENCE

COMMENT JOURNAL  
FEATURES  
source

```
1.193050
  /organism="Ralstonia solanacearum"
  /mol_type="genomic DNA"
  /strain="GMI1000"
  /db_xref="taxon:305"
```

## gene

## CDS

```
505..771
/gene="RSC0962"
/note="synonym: RS04402"
505..771
/gene="RSC0962"
/functions="elements of external origin; phage-related
functions and prophages"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE BACTERIOPHAGE DNA-BINDING TRANSCRIPTION
REGULATOR PROTEIN"
/protein_id="CAD14664.1"
/db_xref="GI:17427975"
/db_xref="UniProt/TREMBL:O8Y0T1"
/translacion="MKYORAHAVAAACSCPTPAPAVSEATARFLRLPEVMSACGUS
RSSIVGVSCGEPFQPVRLPSSVAVWDSIRQWIAIRAROS"
789..1517
/gene="RSC0963"
/note="synonym: RS04401"
789..1517
/gene="RSC0963"
/functions="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHEITICAL PROTEIN"
/protein_id="CAD14665.1"
/db_xref="GI:17427976"
/db_xref="UniProt/TREMBL:O8Y0T0"
/translacion="MKONTVAALPENTLOALTRHONARKGFAPARDEADRIAAMOK
HRTIADSETBAQOAROEAKIMRSTTAQDRLDKAKERAAVATMEDYRAIYAE
EAHVADAKIGAVAKLDENAAVRYLVQYADTMSKAKOMLVPLRAIYAVERTISA
TRNAAAYENNESADALALRMVGVIKQAFARFPAAGAVIQAALRPADLRFIFP
SPGAIHVRNTLRKAAQAQCGRAA"
1811..2032
/gene="RSC0964"
/note="synonym: RS04403"
1811..2032
/gene="RSC0964"
/functions="elements of external origin; phage-related
functions and prophages"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE BACTERIOPHAGE TRANSCRIPTIONAL
ACTIVATOR-RELATED TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD14666.1"
/db_xref="GI:17427977"
/db_xref="GOA:Q810S9"
/db_xref="UniProt/TREMBL:O8Y0S9"
/translacion="MNCPCGSAVATIRTSVPVIRITRELYCQCSNVLGHTFVSLVEV
VRLSPSTPDPPVAROLAGRSEPVAAV"
2400..2576
/gene="RSC0965"
/note="synonym: RS04404"
2400..2576
/gene="RSC0965"
/functions="miscellaneous; unknown"
```

## gene

## CDS

```
505..771
/gene="RSC0962"
/note="synonym: RS04402"
505..771
/gene="RSC0962"
/functions="elements of external origin; phage-related
functions and prophages"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE BACTERIOPHAGE DNA-BINDING TRANSCRIPTION
REGULATOR PROTEIN"
/protein_id="CAD14664.1"
/db_xref="GI:17427975"
/db_xref="UniProt/TREMBL:O8Y0T1"
/translacion="MKYORAHAVAAACSCPTPAPAVSEATARFLRLPEVMSACGUS
RSSIVGVSCGEPFQPVRLPSSVAVWDSIRQWIAIRAROS"
789..1517
/gene="RSC0963"
/note="synonym: RS04401"
789..1517
/gene="RSC0963"
/functions="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHEITICAL PROTEIN"
/protein_id="CAD14665.1"
/db_xref="GI:17427976"
/db_xref="UniProt/TREMBL:O8Y0T0"
/translacion="MKONTVAALPENTLOALTRHONARKGFAPARDEADRIAAMOK
HRTIADSETBAQOAROEAKIMRSTTAQDRLDKAKERAAVATMEDYRAIYAE
EAHVADAKIGAVAKLDENAAVRYLVQYADTMSKAKOMLVPLRAIYAVERTISA
TRNAAAYENNESADALALRMVGVIKQAFARFPAAGAVIQAALRPADLRFIFP
SPGAIHVRNTLRKAAQAQCGRAA"
1811..2032
/gene="RSC0964"
/note="synonym: RS04403"
1811..2032
/gene="RSC0964"
/functions="elements of external origin; phage-related
functions and prophages"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE BACTERIOPHAGE TRANSCRIPTIONAL
ACTIVATOR-RELATED TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD14666.1"
/db_xref="GI:17427977"
/db_xref="GOA:Q810S9"
/db_xref="UniProt/TREMBL:O8Y0S9"
/translacion="MNCPCGSAVATIRTSVPVIRITRELYCQCSNVLGHTFVSLVEV
VRLSPSTPDPPVAROLAGRSEPVAAV"
2400..2576
/gene="RSC0965"
/note="synonym: RS04404"
2400..2576
/gene="RSC0965"
/functions="miscellaneous; unknown"
```

## gene

## CDS

```
505..771
/gene="RSC0962"
/note="synonym: RS04402"
505..771
/gene="RSC0962"
/functions="elements of external origin; phage-related
functions and prophages"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE BACTERIOPHAGE TRANSCRIPTIONAL
ACTIVATOR-RELATED TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD14666.1"
/db_xref="GI:17427977"
/db_xref="GOA:Q810S9"
/db_xref="UniProt/TREMBL:O8Y0S9"
/translacion="MNCPCGSAVATIRTSVPVIRITRELYCQCSNVLGHTFVSLVEV
VRLSPSTPDPPVAROLAGRSEPVAAV"
2400..2576
/gene="RSC0965"
/note="synonym: RS04404"
2400..2576
/gene="RSC0965"
/functions="miscellaneous; unknown"
```

## gene

## CDS

```
1.193050
  /organism="Ralstonia solanacearum"
  /mol_type="genomic DNA"
  /strain="GMI1000"
  /db_xref="taxon:305"
```

```
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD14667.1"
/db_xref="GI:17427978"
/db_xref="UniProt/TREMBL:O8Y0S8"
/translation="MLDAPARPEQPAFPQIIAIVTALRDVAAPTDPAASLDVAGAL
LAVAALADARRRHG"
2569..2892
/gene="RSC0966"
/note="synonym: RS04400"
2569..2892
/gene="RSC0966"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD14668.1"
/db_xref="GI:17427979"
/db_xref="UniProt/TREMBL:O8Y0S7"
/translation="MKRNHPHWALEKQSLTQIAQOAPQTPPAATLARLETVS
ASMRDATERAGQLLCVSALDGLICLDLHAQTAPGYRSLHCLLVKQQLDGAIGT
VTENL"
2903..5704
/gene="RSC0967"
/note="synonym: RS04399"
2903..5704
/gene="RSC0967"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD14669.1"
/db_xref="GI:17427980"
/db_xref="GOA:O8Y0S6"
/db_xref="UniProt/TREMBL:O8Y0S6"
/translation="MHTQSPALRTWTFDQDKRIDLDHLAERLDLRSGAKGNYSPI
HTDRASALSIPIFERGAMKDMSTEGSSCIDYOHCMQDVATPMEAAXKILGOWYGI PMP
AAPASALLARKSTERYIARAQANPEPAVAYLAGGVDEAVSRRAIQGALGNMANS
PVPAGEAGHGPAAPAFIVRAMDTAVAVADLVADPAINGVAKIQCGDKLGHMTS
DARRLHAHTVYIVESPINALVECHLPEGTAAFLRTANVDKI DMWTFLEKRVVI
ALDHDTPVNERGTMRPGLAANKLSALTAADIGAMLYDMQWEGEDINDVLOAHGA
AELTAMRRLAMLIJGMPGGERFLGSTRVFIYPHDFGIYRFRYKEDFTQYVEKF
DADDEGQTSRSBELGDCAPRVAGLSRLRVOSHLATINGTADSOPEYFQI SAQV
RGATLQREYVGNDKLYNLEWRGKFGHIMPSQFARMNVYLERAADLGARDVYVGL
AWRGELALLESDCYVEPQKQCLYNNRFPRTSQARAVIDAIVQATFKGNAATA
LVWALGTHLKTIVGFYPHLOMAEKAGKAGKLSLQATASQVLSGMLKTDHRRRA
SVSWTSHPVGMDSEFKLPAVLSIEDGLQSTYRFEFTVGAALTPYIMCAPVLAGE
EVDVESLQSKICRTSLSDVKQAILPRNLPOPPVAMVLOFLASOOPERIDHAAFD
VLSRRAETNDATAPRWENYAAITMTAALLAEPAQIDVEQGFI DHIILVEMNAHIA
ETDGTRLPVWYIMEILSELSDAGREPHPCWPTTIQTSGREMALFLRSHVNDHSTA
MHLRGKFDALPIKTARVFKQQLQSGVVPQHGKRLDEVKRIIGARRCAHMAISLPR
LEORGLYASPALAAYOMAS"
complement (6298..7473)
/gene="RSC0968"
/note="synonym: RS04398"
complement (6298..7473)
```

```
/gene="RSC0968"
/functions="elements of external origin; transposon-related
functions"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE CP4-LIKE INTEGRASE PROTEIN"

Query Match 78.3%; Score 18.8; DB 1; Length 193050;
Best Local Similarity 70.8%; Pred. No. 3.3e+03;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCCGSGTGGCCGSGTSACSGSAC 24
Db 78885 GCCCTGCTGCGCGGTCAAGCCGAC 78908
```

Search completed: July 20, 2005, 16:45:50  
Job time : 1178.57 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 16:46:04 ; Search time 429.857 Seconds

(without alignments)  
265.662 Million cell updates/sec

Title: US-10-017-471B-1

Perfect score: 18

Sequence: 1 gaccacgctccggagcatg 18

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 7173243 seqs, 3172129809 residues

Total number of hits satisfying chosen parameters: 14346486

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 300 summaries

Database : Published Applications\_NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PTCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PTCT\_NEW\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US10J\_PUBCOMB.seq:\*  
23: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq:\*  
24: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.2	95.6	18	15	US-10-017-471A-1
2	17.2	95.6	891	15	US-10-156-761-3511
3	17.2	95.6	9025608	15	US-10-156-761-1
4	16	88.9	4346	15	US-10-017-471A-19
5	15.6	86.7	513	10	US-09-975-719-155
6	15.6	86.7	612	22	US-10-617-320-417
7	15.6	86.7	762	10	US-09-975-719-153

C 8	15.6	86.7	1080	13	US-10-027-632-100941	Sequence 100941,
C 9	15.6	86.7	1080	13	US-10-027-632-100942	Sequence 100942,
C 10	15.6	86.7	1080	13	US-10-027-632-100943	Sequence 100943,
C 11	15.6	86.7	1080	17	US-10-027-632-100941	Sequence 100941,
C 12	15.6	86.7	1080	17	US-10-027-632-100942	Sequence 100942,
C 13	15.6	86.7	1080	17	US-10-027-632-100943	Sequence 100943,
C 14	15.6	86.7	1080	21	US-10-472-928-473	Sequence 473, App
C 15	15.6	86.7	2048	10	US-09-975-719-136	Sequence 136, App
C 16	15.6	86.7	2395	8	US-08-961-527-205	Sequence 205, App
C 17	15.6	86.7	713059	13	US-10-158-844-205	Sequence 205, App
C 18	15.6	86.7	713059	17	US-10-027-632-174581	Sequence 174581,
C 19	15.6	86.7	713059	13	US-10-027-632-174581	Sequence 174581,
C 20	15.6	86.7	2162598	21	US-10-472-928-4979	Sequence 4979, App
C 21	15.2	84.4	330	15	US-10-156-761-6136	Sequence 6136, App
C 22	15.2	84.4	110021	21	US-10-461-862-83	Sequence 83, App
C 23	14.6	81.1	209	9	US-09-294-0938-4373	Sequence 4374, App
C 24	14.6	81.1	271	20	US-10-425-115-76979	Sequence 76979, A
C 25	14.6	81.1	521	20	US-10-425-115-131598	Sequence 131598,
C 26	14.6	81.1	521	20	US-10-856-499-238	Sequence 238, App
C 27	14.6	81.1	521	20	US-10-856-499-1210	Sequence 1210, App
C 28	14.6	81.1	558	15	US-10-259-165-501	Sequence 501, App
C 29	14.6	81.1	735	19	US-10-437-963-73433	Sequence 73433, A
C 30	14.6	81.1	1041	17	US-10-107-431-2	Sequence 2, App
C 31	14.6	81.1	1068	17	US-10-084-846A-53	Sequence 53, App
C 32	14.6	81.1	1092	17	US-10-369-493-40179	Sequence 40179, A
C 33	14.6	81.1	1209	17	US-10-369-493-24261	Sequence 24261, A
C 34	14.6	81.1	1266	15	US-10-156-761-2078	Sequence 2078, App
C 35	14.6	81.1	1482	15	US-10-156-761-4032	Sequence 4032, App
C 36	14.6	81.1	1674	17	US-10-369-493-35833	Sequence 35833, A
C 37	14.6	81.1	1707	17	US-10-369-493-24890	Sequence 24890, A
C 38	14.6	81.1	1749	9	US-09-821-167-13	Sequence 13, App
C 39	14.6	81.1	254	17	US-10-108-260A-1946	Sequence 1946, App
C 40	14.6	81.1	3011	19	US-10-437-963-40436	Sequence 40436, A
C 41	14.6	81.1	3799	17	US-10-369-493-26821	Sequence 26821, A
C 42	14.6	81.1	5874	17	US-10-365-747-119	Sequence 119, App
C 43	14.6	81.1	7564	19	US-10-476-397-15	Sequence 15, App
C 44	14.6	81.1	11188	9	US-09-821-167-1	Sequence 1, App
C 45	14.6	81.1	14061	17	US-10-093-463-71	Sequence 71, App
C 46	14.6	81.1	14109	17	US-10-093-463-71	Sequence 71, App
C 47	14.6	81.1	17083	19	US-10-475-970-4	Sequence 4, App
C 48	14.6	81.1	20489	20	US-10-723-860-5701	Sequence 5701, App
C 49	14.6	81.1	23907	13	US-10-077-130-6	Sequence 6, App
C 50	14.6	81.1	24120	13	US-10-077-130-4	Sequence 4, App
C 51	14.6	81.1	33825	21	US-10-866-089-11	Sequence 1, App
C 52	14.6	81.1	33825	21	US-10-866-089-12	Sequence 12, App
C 53	14.6	81.1	43865	19	US-10-322-281-226	Sequence 226, App
C 54	14.6	81.1	45055	17	US-10-107-431-277	Sequence 277, App
C 55	14.6	81.1	59816	17	US-10-084-846A-2	Sequence 2, App
C 56	14.6	81.1	59816	17	US-10-084-846A-1	Sequence 1, App
C 57	14.6	81.1	71222	13	US-10-087-193-1942	Sequence 1942, App
C 58	14.6	81.1	2356646	19	US-10-087-193-1942	Sequence 1, App
C 59	14.6	81.1	2356646	19	US-10-470-565-1	Sequence 1, App
C 60	14.6	81.1	9025608	15	US-10-156-761-1	Sequence 1, App
C 61	14.4	80.0	144	16	US-10-012-952A-33	Sequence 33, App
C 62	14.4	80.0	455	10	US-09-918-995-2309	Sequence 2309, App
C 63	14.4	80.0	4921	19	US-10-437-963-82224	Sequence 82224, A
C 64	14.4	80.0	1116	17	US-10-369-493-39328	Sequence 39328, A
C 65	14.4	80.0	1116	17	US-10-369-493-39703	Sequence 39703, A
C 66	14.4	80.0	1116	17	US-10-369-493-40056	Sequence 40056, A
C 67	14.4	80.0	1125	18	US-10-425-115-7140	Sequence 7140, App
C 68	14.4	80.0	1185	17	US-10-369-493-46060	Sequence 46060, A
C 69	14.4	80.0	1868	18	US-10-425-115-5724	Sequence 5724, App
C 70	14.4	80.0	2237	19	US-10-437-963-84314	Sequence 84314, A
C 71	14.4	80.0	4017	20	US-10-425-115-40887	Sequence 40887, A
C 72	14.4	80.0	4017	19	US-10-437-963-19901	Sequence 19901, A
C 73	14.4	80.0	4323	19	US-10-437-963-18997	Sequence 18997, A
C 74	14.4	80.0	3853122	21	US-10-741-600-17796	Sequence 17796, A
C 75	14.4	80.0	653122	13	US-10-087-193-226	Sequence 226, App
C 76	14.4	80.0	653122	20	US-10-087-193-226	Sequence 226, App
C 77	14.2	78.9	233	10	US-10-425-115-84144	Sequence 84144, A
C 78	14.2	78.9	471	9	US-09-934-903-9	Sequence 9, App
C 79	14.2	78.9	471	9	US-09-934-868-65	Sequence 65, App
C 80	14.2	78.9	471	10	US-09-941-947A-13	Sequence 13, App

81	14.2	78.9	471	18	US-10-700-003-9	Sequence 9, Appli	C 154	14	77.8	564	20	US-10-425-115-138353	Sequence 138353,
82	14.2	78.9	471	18	US-10-363-567-13	Sequence 11, Appl	155	14	77.8	565	13	US-10-027-632-137998	Sequence 137998,
83	14.2	78.9	471	22	US-10-701-200-65	Sequence 65, Appl	156	14	77.8	565	17	US-10-027-632-137998	Sequence 137998,
84	14.2	78.9	474	15	US-10-156-761-4846	Sequence 4846, A	157	14	77.8	573	20	US-10-424-599-8518	Sequence 8518, Ap
85	14.2	78.9	575	20	US-10-425-115-37988	Sequence 37988, A	158	14	77.8	574	18	US-10-425-115-118455	Sequence 118455,
C 86	14.2	78.9	579	15	US-10-259-165-701	Sequence 701, App	159	14	77.8	588	15	US-10-156-761-3225	Sequence 3225, Ap
C 87	14.2	78.9	610	19	US-10-767-701-8566	Sequence 8566, Ap	160	14	77.8	604	19	US-10-767-701-6259	Sequence 6259, App
88	14.2	78.9	610	15	US-10-156-761-180	Sequence 180, App	C 161	14	77.8	618	9	US-09-770-149-008	Sequence 808, App
89	14.2	78.9	741	15	US-10-156-761-6450	Sequence 6450, App	C 162	14	77.8	637	13	US-10-027-632-220440	Sequence 220440,
90	14.2	78.9	796	15	US-10-062-831-44	Sequence 44, Appl	C 163	14	77.8	637	13	US-10-027-632-220441	Sequence 220441,
91	14.2	78.9	796	16	US-10-062-599-44	Sequence 44, Appl	C 164	14	77.8	637	13	US-10-027-632-220442	Sequence 220442,
92	14.2	78.9	897	17	US-10-369-483-39451	Sequence 39451, A	C 165	14	77.8	637	17	US-10-027-632-220443	Sequence 220443,
93	14.2	78.9	897	17	US-10-369-483-39835	Sequence 39835, A	C 166	14	77.8	637	17	US-10-027-632-220440	Sequence 220440,
94	14.2	78.9	909	17	US-10-369-483-39835	Sequence 39835, A	C 167	14	77.8	637	17	US-10-027-632-220441	Sequence 220441,
95	14.2	78.9	909	17	US-10-369-483-39835	Sequence 39835, A	C 168	14	77.8	637	17	US-10-027-632-220442	Sequence 220442,
96	14.2	78.9	1179	21	US-10-200-545-18	Sequence 18, Appl	C 169	14	77.8	637	13	US-10-027-632-131009	Sequence 131009,
C 97	14.2	78.9	1295	14	US-10-281-024-2	Sequence 2, Appli	C 170	14	77.8	660	17	US-10-027-632-131009	Sequence 131009,
98	14.2	78.9	1504	20	US-10-425-115-139009	Sequence 139009,	171	14	77.8	661	19	US-10-767-701-3355	Sequence 3355, Ap
99	14.2	78.9	1836	17	US-10-369-483-39783	Sequence 39783, A	C 172	14	77.8	666	9	US-09-770-149-445	Sequence 445, App
100	14.2	78.9	2035	17	US-10-369-483-39026	Sequence 39026, A	C 173	14	77.8	669	19	US-10-437-963-54737	Sequence 54737, A
101	14.2	78.9	2035	17	US-10-369-483-39026	Sequence 39026, A	C 174	14	77.8	671	10	US-09-952-060-13	Sequence 13, Appl
102	14.2	78.9	2078	19	US-10-437-963-71557	Sequence 71557, A	C 175	14	77.8	671	14	US-10-149-640-5	Sequence 5, Appl
103	14.2	78.9	2343	20	US-10-437-963-71557	Sequence 71557, A	C 176	14	77.8	671	14	US-10-380-641-13	Sequence 13, Appl
C 104	14.2	78.9	2343	20	US-10-437-963-71557	Sequence 71557, A	C 177	14	77.8	671	19	US-10-380-641-13	Sequence 13, Appl
C 105	14.2	78.9	2627	17	US-10-369-483-37750	Sequence 37750, A	C 178	14	77.8	671	21	US-10-636-730-13	Sequence 730,13
C 106	14.2	78.9	4122	19	US-10-437-963-56914	Sequence 56914, A	C 179	14	77.8	671	15	US-10-156-761-7519	Sequence 7519, Ap
C 107	14.2	78.9	4122	19	US-10-437-963-56914	Sequence 56914, A	C 180	14	77.8	684	18	US-10-425-114-31951	Sequence 31951, A
C 108	14.2	78.9	60	10	US-09-804-980-200	Sequence 452265,	C 181	14	77.8	705	20	US-10-425-115-75219	Sequence 75219, A
C 109	14.2	78.9	60	10	US-09-804-980-200	Sequence 452265,	C 182	14	77.8	728	20	US-10-425-115-11625	Sequence 11625,
C 110	14.2	78.9	139	14	US-10-149-640-28	Sequence 28, App	C 183	14	77.8	740	19	US-10-437-963-93255	Sequence 93255, A
C 111	14.2	78.9	255	17	US-10-084-846A-13	Sequence 28, Appl	C 184	14	77.8	763	17	US-10-139-794-210	Sequence 210, App
C 112	14.2	78.9	256	17	US-10-084-846A-13	Sequence 13, Appl	C 185	14	77.8	768	20	US-10-425-115-9753	Sequence 9753, App
C 113	14.2	78.9	268	9	US-09-922-217-723	Sequence 110668,	C 186	14	77.8	789	17	US-10-369-493-35141	Sequence 35141, A
C 114	14.2	78.9	268	9	US-09-922-217-723	Sequence 723, App	C 187	14	77.8	789	17	US-10-369-493-38322	Sequence 38322, A
C 115	14.2	78.9	268	13	US-10-025-380-773	Sequence 723, App	C 188	14	77.8	789	17	US-10-369-493-38784	Sequence 38784, A
C 116	14.2	78.9	291	10	US-09-804-980-174	Sequence 723, App	C 189	14	77.8	801	20	US-10-425-115-15290	Sequence 15290,
C 117	14.2	78.9	291	10	US-09-804-980-174	Sequence 723, App	C 190	14	77.8	801	20	US-10-425-115-85980	Sequence 85980, A
C 118	14.2	78.9	299	20	US-10-425-115-72578	Sequence 198, App	C 191	14	77.8	836	20	US-10-425-115-85980	Sequence 85980, A
C 119	14.2	78.9	303	17	US-10-282-122A-19843	Sequence 72578, A	C 192	14	77.8	840	17	US-10-369-493-35141	Sequence 35141, A
C 120	14.2	78.9	311	19	US-10-767-701-17402	Sequence 19843, A	C 193	14	77.8	868	20	US-10-425-115-98251	Sequence 98251, A
C 121	14.2	78.9	312	17	US-10-242-535A-47073	Sequence 17402, A	C 194	14	77.8	876	19	US-10-767-701-10747	Sequence 10747, A
C 122	14.2	78.9	312	17	US-10-242-535A-47073	Sequence 47073, A	C 195	14	77.8	895	18	US-10-424-599-37600	Sequence 37600, A
C 123	14.2	78.9	322	18	US-10-424-599-102046	Sequence 47073, A	C 196	14	77.8	895	18	US-10-424-599-37600	Sequence 37600, A
C 124	14.2	78.9	322	18	US-10-424-599-102046	Sequence 102046,	C 197	14	77.8	913	18	US-10-425-114-22860	Sequence 22860, A
C 125	14.2	78.9	336	19	US-10-437-963-84888	Sequence 188, App	C 198	14	77.8	918	18	US-10-425-114-28280	Sequence 28280, A
C 126	14.2	78.9	336	19	US-10-437-963-84888	Sequence 84888, A	C 199	14	77.8	928	18	US-10-425-114-28312	Sequence 28312, A
C 127	14.2	78.9	343	20	US-10-425-115-164181	Sequence 164181,	C 200	14	77.8	933	18	US-10-425-114-28318	Sequence 28318, A
C 128	14.2	78.9	343	20	US-10-425-115-164181	Sequence 164181,	C 201	14	77.8	942	18	US-10-424-599-91985	Sequence 91985, A
C 129	14.2	78.9	348	10	US-09-803-719-92	Sequence 36422, A	C 202	14	77.8	945	18	US-10-425-114-28319	Sequence 28319, A
C 130	14.2	78.9	356	11	US-09-864-408A-3221	Sequence 92, Appl	C 203	14	77.8	948	20	US-10-425-115-70627	Sequence 70627, A
C 131	14.2	78.9	356	11	US-09-864-408A-3221	Sequence 64, Appl	C 204	14	77.8	949	18	US-10-425-114-28554	Sequence 28554, A
C 132	14.2	78.9	364	20	US-10-425-115-89605	Sequence 3221, App	C 205	14	77.8	951	18	US-10-425-114-28178	Sequence 28178, A
C 133	14.2	78.9	390	20	US-10-425-115-20563	Sequence 89605, A	C 206	14	77.8	951	18	US-10-425-114-28178	Sequence 28178, A
C 134	14.2	78.9	413	10	US-09-918-995-4176	Sequence 20563, A	C 207	14	77.8	953	18	US-10-425-114-28178	Sequence 28178, A
C 135	14.2	78.9	416	18	US-10-424-599-36816	Sequence 4176, Ap	C 208	14	77.8	958	18	US-10-425-114-28133	Sequence 28133, A
C 136	14.2	78.9	428	10	US-09-918-995-16713	Sequence 36816, Ap	C 209	14	77.8	971	18	US-10-425-114-15752	Sequence 15752, A
C 137	14.2	78.9	428	19	US-10-437-963-14432	Sequence 16713, A	C 210	14	77.8	979	18	US-10-240-140A-117	Sequence 117, App
C 138	14.2	78.9	449	20	US-10-425-115-67593	Sequence 14432, A	C 211	14	77.8	1041	17	US-10-107-431-12	Sequence 107, Appl
C 139	14.2	78.9	451	9	US-09-960-352-1342	Sequence 67593, A	C 212	14	77.8	1041	17	US-10-107-431-12	Sequence 107, Appl
C 140	14.2	78.9	458	20	US-10-425-115-38190	Sequence 1342, App	C 213	14	77.8	1047	17	US-10-084-646A-9	Sequence 8, Appli
C 141	14.2	78.9	467	10	US-09-918-995-30974	Sequence 38190, A	C 214	14	77.8	1047	17	US-10-084-646A-9	Sequence 8, Appli
C 142	14.2	78.9	469	10	US-09-918-995-32468	Sequence 30974, A	C 215	14	77.8	1049	17	US-10-487-901-6180	Sequence 6180, App
C 143	14.2	78.9	469	20	US-10-425-115-18769	Sequence 32468, A	C 216	14	77.8	1124	21	US-10-487-901-6180	Sequence 6180, App
C 144	14.2	78.9	477	17	US-10-282-122A-6982	Sequence 18769, A	C 217	14	77.8	1124	21	US-10-369-493-47075	Sequence 47075, A
C 145	14.2	78.9	479	20	US-10-425-115-20558	Sequence 6982, Ap	C 218	14	77.8	1156	20	US-10-425-115-85991	Sequence 85991, A
C 146	14.2	78.9	482	13	US-10-027-632-278807	Sequence 20558, A	C 219	14	77.8	1156	20	US-10-425-115-85991	Sequence 85991, A
C 147	14.2	78.9	482	13	US-10-027-632-278807	Sequence 278807,	C 220	14	77.8	1167	10	US-09-933-167-107	Sequence 107, App
C 148	14.2	78.9	492	19	US-10-425-115-86991	Sequence 278807,	C 221	14	77.8	1167	10	US-10-004-660-107	Sequence 107, App
C 149	14.2	78.9	538	20	US-10-425-115-28191	Sequence 86991, Ap	C 222	14	77.8	1167	10	US-10-004-660-107	Sequence 107, App
C 150	14.2	78.9	539	20	US-10-425-115-152949	Sequence 28191, A	C 223	14	77.8	1167	10	US-10-004-660-107	Sequence 107, App
C 151	14.2	78.9	540	9	US-09-854-133-658	Sequence 152949,	C 224	14	77.8	1167	10	US-10-004-660-107	Sequence 107, App
C 152	14.2	78.9	540	15	US-10-144-649A-658	Sequence 658, App	C 225	14	77.8	1167	10	US-10-004-660-107	Sequence 107, App
C 153	14.2	78.9	550	20	US-10-425-115-20560	Sequence 20560, A	C 226	14	77.8	1179	16	US-10-091-007-823	Sequence 223, App

C 227	14	77.8	1182	17	US-10-369-493-24247	Sequence 24217, A
C 228	14	77.8	1200	20	US-10-425-115-20566	Sequence 20566, A
C 229	14	77.8	1218	16	US-10-174-209-25	Sequence 25, Appl
C 230	14	77.8	1227	15	US-10-156-761-3900	Sequence 3900, A
C 231	14	77.8	1235	19	US-10-437-963-86392	Sequence 86392, A
C 232	14	77.8	1266	17	US-10-369-493-45504	Sequence 45504, A
C 233	14	77.8	1295	19	US-10-437-963-9754	Sequence 9754, Ap
C 234	14	77.8	1308	17	US-10-282-122A-33059	Sequence 33059, A
C 235	14	77.8	1319	19	US-10-437-963-68782	Sequence 68782, A
C 236	14	77.8	1326	20	US-10-425-115-175783	Sequence 175783, A
C 237	14	77.8	1335	18	US-10-425-114-14047	Sequence 14047, A
C 238	14	77.8	1341	18	US-10-425-114-32758	Sequence 32758, A
C 239	14	77.8	1356	9	US-09-938-330-1	Sequence 1, Appl
C 240	14	77.8	1359	19	US-10-437-963-20972	Sequence 20972, A
C 241	14	77.8	1362	18	US-10-425-114-35232	Sequence 35232, A
C 242	14	77.8	1377	20	US-10-437-963-64659	Sequence 64659, A
C 243	14	77.8	1377	20	US-10-411-910A-266	Sequence 266, App
C 244	14	77.8	1381	20	US-10-425-115-129730	Sequence 129730, A
C 245	14	77.8	1386	17	US-10-369-493-31705	Sequence 31705, A
C 246	14	77.8	1389	17	US-10-417-700A-110	Sequence 110, App
C 247	14	77.8	1392	18	US-10-417-700A-104	Sequence 104, App
C 248	14	77.8	1395	18	US-10-417-700A-108	Sequence 108, App
C 249	14	77.8	1410	18	US-10-417-700A-106	Sequence 106, App
C 250	14	77.8	1412	20	US-10-425-115-16343	Sequence 16343, A
C 251	14	77.8	1440	9	US-09-712-363-94	Sequence 94, Appl
C 252	14	77.8	1440	17	US-10-282-122A-28525	Sequence 28525, A
C 253	14	77.8	1452	17	US-10-282-122A-26564	Sequence 26564, A
C 254	14	77.8	1461	9	US-09-938-330-5	Sequence 5, Appl
C 255	14	77.8	1476	17	US-10-282-122A-28799	Sequence 28799, A
C 256	14	77.8	1479	15	US-10-156-761-1451	Sequence 1451, Ap
C 257	14	77.8	1482	15	US-10-282-122A-11680	Sequence 11680, A
C 258	14	77.8	1488	21	US-10-490-913-3	Sequence 3, Appl
C 259	14	77.8	1488	21	US-10-887-553A-756	Sequence 756, Appl
C 260	14	77.8	1488	21	US-10-956-157-1639	Sequence 1639, App
C 261	14	77.8	1488	21	US-10-956-157-1639	Sequence 1639, App
C 262	14	77.8	1491	19	US-10-437-963-55898	Sequence 55898, A
C 263	14	77.8	1494	19	US-10-282-122A-28783	Sequence 28783, A
C 264	14	77.8	1503	19	US-10-437-963-13836	Sequence 13836, A
C 265	14	77.8	1503	21	US-10-965-898-120	Sequence 120, App
C 266	14	77.8	1524	16	US-10-204-220-5	Sequence 5, Appl
C 267	14	77.8	1530	19	US-10-437-963-47467	Sequence 47467, A
C 268	14	77.8	1559	17	US-10-282-122A-19326	Sequence 19326, A
C 269	14	77.8	1563	21	US-10-956-157-4211	Sequence 4211, Ap
C 270	14	77.8	1593	17	US-10-282-122A-25611	Sequence 25611, A
C 271	14	77.8	1597	19	US-10-437-963-20809	Sequence 20809, A
C 272	14	77.8	1616	20	US-10-425-115-180527	Sequence 180527, A
C 273	14	77.8	1660	17	US-10-415-378-27	Sequence 27, Appl
C 274	14	77.8	1680	9	US-09-974-300-2326	Sequence 2326, Ap
C 275	14	77.8	1686	18	US-10-425-114-32782	Sequence 32782, A
C 276	14	77.8	1704	15	US-10-156-761-4987	Sequence 4987, A
C 277	14	77.8	1734	15	US-10-156-761-29	Sequence 29, Appl
C 278	14	77.8	1740	20	US-10-425-115-85990	Sequence 85990, A
C 279	14	77.8	1811	9	US-09-822-830A-617	Sequence 617, App
C 280	14	77.8	1832	18	US-10-425-114-28132	Sequence 28132, A
C 281	14	77.8	1838	19	US-10-437-963-88815	Sequence 88815, A
C 282	14	77.8	1857	17	US-10-282-122A-15065	Sequence 15065, A
C 283	14	77.8	1866	14	US-10-166-087-13	Sequence 13, Appl
C 284	14	77.8	1879	18	US-10-425-114-32261	Sequence 32261, A
C 285	14	77.8	1879	18	US-10-641-643-742	Sequence 742, App
C 286	14	77.8	1882	17	US-10-108-260A-1035	Sequence 1035, Ap
C 287	14	77.8	1963	20	US-10-739-930-4011	Sequence 4011, Ap
C 288	14	77.8	2055	18	US-10-425-114-29541	Sequence 29541, A
C 289	14	77.8	2058	20	US-10-425-115-85988	Sequence 85988, A
C 290	14	77.8	2103	15	US-10-156-761-4481	Sequence 4481, Ap
C 291	14	77.8	2103	17	US-10-282-122A-25741	Sequence 25741, A
C 292	14	77.8	2106	20	US-10-425-115-46880	Sequence 46880, A
C 293	14	77.8	2166	17	US-10-330-011-127	Sequence 127, App
C 294	14	77.8	2176	17	US-10-264-237-994	Sequence 994, App
C 295	14	77.8	2187	18	US-10-425-114-30640	Sequence 30640, A
C 296	14	77.8	2212	20	US-10-424-599-72496	Sequence 72496, A
C 297	14	77.8	2226	20	US-10-425-115-47014	Sequence 47014, A
C 298	14	77.8	2334	17	US-10-260-238-1133	Sequence 1133, Ap
C 299	14	77.8	2367	19	US-10-437-963-43182	Sequence 43182, A

300 14 77.8 2394 17 US-10-369-493-26878 Sequence 26878, A

## ALIGNMENTS

```

RESULT 1
US-10-017-471A-1
; Sequence 1, Application US/10017471A
; Publication No. US20030124644A1
; GENERAL INFORMATION:
; APPLICANT: Takano, Eriko
; APPLICANT: Bibb, Mervyn
; TITLE OF INVENTION: Antibiotic Production
; FILE REFERENCE: 0380-P0329US1
; CURRENT APPLICATION NUMBER: US/10/017,471A
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,561
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-017-471A-1

```

```

Query Match          95.6%; Score 17.2; DB 15; Length 18;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GACCACTGCCGCGCATG 18
Db      1 GACCACTGCCGCGCATG 18

```

```

RESULT 2
US-10-156-761-3511
; Sequence 3511, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3511
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (1)..(891)
US-10-156-761-3511

```

```

Query Match          95.6%; Score 17.2; DB 15; Length 891;
Best Local Similarity 88.9%; Pred. No. 41;
Matches 16; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 GACCACGTSCCGGCATG 18
          |||||:|:|||||
Db      160 GACCACGTCCCCGGCATG 177

```

```

RESULT 3
US-10-156-761-1/c
Sequence 1, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIEDA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JIN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OR INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermectilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

	Query Match	Similarity	Score	DB	Length
Best Local	Similarity	95.6%	17.2	DB 15	9025608
Matches	16	Conservative	2	Mismatches	0
				Indels	0
				Gaps	0

```

1  RESULT 4
2  US-10-017-471A-19/c
3  Sequence 19, Application US/10017471A
4  Publication No. US20030124644A1
5  GENERAL INFORMATION:
6  APPLICANT: Takano, Eriko
7  APPLICANT: Bibb, Merynn
8  TITLE OF INVENTION: Antibiotic Production
9  FILE REFERENCE: 0380-P02329US1
10 CURRENT APPLICATION NUMBER: US/10/017,471A
11 CURRENT FILING DATE: 2001-10-23
12 PRIOR APPLICATION NUMBER: 2000-10-23
13 PRIOR FILING DATE: 2000-10-23
14 NUMBER OF SEQ ID NOS: 19
15 SOFTWARE: PatentIn Ver. 2.1
16 SEQ ID NO 19
17 LENGTH: 4346
18 TYPE: DNA
19 ORGANISM: Streptomyces coelicolor
20 US-10-017-471A-19

```

Query Match	Score	DB	Length
Best Local Similarity	88.9%	Pred. No. 1	2e+02
Matches	16	Conservative	1, Mismatches 1, Indels 0, Gaps 0

  

QY	1	GACCACTGSCGGCATG	18
DB	1453	GACCACTGACGGGCATG	1436

```

RESULT 5
US-09-975-719-155
; Sequence 155, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Anshubel, Frederick M.
; APPLICANT: Rahme, Lawrence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 155
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-155

```

Query Match	86.7%	Score 15.6	DB 10	Length 513
Best Local Similarity	83.3%	Pred. No. 2.9e+02		
Matches	15	Conservative	2	Mismatches 1
				Indels 0
				Gaps 0
QY	1	GACCACTGCTCCGGCATG	18	
Db	133	GACCACTGCTCCGGCATG	150	

RESULT 6  
 US-10-617-320-417/c  
 ; Sequence 417, Application US/10617320  
 ; Publication No. US20050136404A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
 ; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
 ; THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 5206  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; City: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02354  
 ;  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: <Unknown>  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: <Unknown>  
 ;  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/617,320  
 ; FILING DATE: 10-Jul-2003  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,433  
 ; FILING DATE: 30-Jun-1998  
 ; APPLICATION NUMBER: 60/085131  
 ; FILING DATE: May 12, 1998  
 ; APPLICATION NUMBER: 60/051553  
 ; FILING DATE: July 2, 1997  
 ;  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Arinello, Pamela Deneke  
 ; REGISTRATION NUMBER: 40,489  
 ; REFERENCE/DOCKET NUMBER: GTC-011  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (781) 893-5007  
 ;

```

; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 417:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 612 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: (b) LOCATION 1...612
;     SEQUENCE DESCRIPTION: SEQ ID NO: 417:
US-10-617-320-417

Query Match      86.7%; Score 15.6; DB 22; Length 612;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACTSCCGCGCATG 18
Db      160 GACCACTGCCGCGCATG 143

RESULT 7
US-09-975-719-153/c
; Sequence 153, Application US/09975719
; Publication No. US2003002349A1
; GENERAL INFORMATION:
;   APPLICANT: Ausubel, Frederick M.
;   APPLICANT: Rahme, Laurence G.
;   TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
;   FILE REFERENCE: 00786/361003
;   CURRENT APPLICATION NUMBER: US/09/975,719
;   PRIOR FILING DATE: 2001-10-10
;   PRIOR APPLICATION NUMBER: US 09/199,637
;   PRIOR FILING DATE: 1998-11-25
;   PRIOR APPLICATION NUMBER: US 60/066,517
;   PRIOR FILING DATE: 1997-11-25
;   NUMBER OF SEQ ID NOS: 437
;   SOFTWARE: FastSeq for Windows Version 4.0
;   SEQ ID NO 153
;   LENGTH: 762
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-975-719-153

Query Match      86.7%; Score 15.6; DB 10; Length 762;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACTSCCGCGCATG 18
Db      303 GACCACTCCGCGCGCATG 286

RESULT 8
US-10-027-632-100941/c
; Sequence 100941, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSeq for Windows Version 4.0
;   SEQ ID NO 100941
;   LENGTH: 1080
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-100941

Query Match      86.7%; Score 15.6; DB 13; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACTSCCGCGCATG 18
Db      206 GACCACTCCCGCGCAG 189

RESULT 9
US-10-027-632-100942/c
; Sequence 100942, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSeq for Windows Version 4.0
;   SEQ ID NO 100942
;   LENGTH: 1080
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-100942

Query Match      86.7%; Score 15.6; DB 13; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACTSCCGCGCATG 18
Db      206 GACCACTCCCGCGCAG 189

RESULT 10
US-10-027-632-100943/c
```

```

; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100941
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-100941

Query Match      86.7%; Score 15.6; DB 13; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACTSCCGCGCATG 18
Db      206 GACCACTCCCGCGCAG 189

RESULT 10
US-10-027-632-100943/c
```

```
; Sequence 100943, Application US/10027632
; Publication No. US20020198377A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100943
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-100943

Query Match      86.7%; Score 15.6; DB 13; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACGTSCCGGCATG 18
Db      206 GACCACGTCCCGGCAGG 189

RESULT 11
US-10-027-632-100941/c
; Sequence 100941, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100941
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-100941
```

```
Query Match      86.7%; Score 15.6; DB 17; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACGTSCCGGCATG 18
Db      206 GACCACGTCCCGGCAGG 189

RESULT 12
US-10-027-632-100942/c
; Sequence 100942, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100942
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-100942

Query Match      86.7%; Score 15.6; DB 17; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACGTSCCGGCATG 18
Db      206 GACCACGTCCCGGCAGG 189

RESULT 13
US-10-027-632-100943/c
; Sequence 100943, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100943
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-100943
```

PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 100943  
LENGTH: 1080  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-100943

Query Match 86.7%; Score 15.6; DB 17; Length 1080;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAGTSCCGGCGCATG 18  
|||||:|||||  
DB 206 GACCAGTCCCGGCGCAGG 189

RESULT 14  
US-10-472-928-473/c  
Sequence 473, Application US/10472928  
Publication No. US2005020813A1  
GENERAL INFORMATION:  
APPLICANT: CHIRON SPA  
TITLE OF INVENTION: THE INSTITUTE FOR GENOMIC RESEARCH  
FILE REFERENCE: P026926MO  
CURRENT APPLICATION NUMBER: US/10/472,928  
CURRENT FILING DATE: 2003-09-26  
PRIOR APPLICATION NUMBER: GB-0107658.7  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 4979  
SOFTWARE: SeqWin99, version 1.03  
SEQ ID NO 473  
LENGTH: 1188  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
US-10-472-928-473

Query Match 86.7%; Score 15.6; DB 21; Length 1188;  
Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAGTSCCGGCGCATG 18  
|||||:|||||  
DB 739 GACCAGTCCCGGCGCATG 722

RESULT 15  
US-09-975-719-136  
Sequence 136, Application US/09975719  
Publication No. US20030022349A1  
GENERAL INFORMATION:  
APPLICANT: Ausubel, Frederick M.  
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
FILE REFERENCE: 00786/361003  
CURRENT APPLICATION NUMBER: US/09/975,719  
CURRENT FILING DATE: 2001-10-10  
PRIOR APPLICATION NUMBER: US 09/199,637  
PRIOR FILING DATE: 1998-11-25  
PRIOR APPLICATION NUMBER: US 60/066,517  
PRIOR FILING DATE: 1997-11-25  
NUMBER OF SEQ ID NOS: 437  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 136  
LENGTH: 2048  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-975-719-136

Query Match 86.7%; Score 15.6; DB 10; Length 2048;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAGTSCCGGCGCATG 18  
|||||:|||||  
DB 1069 GACCAGTCCCGGCGCATG 1086

RESULT 16  
US-08-961-527-205/c  
Sequence 205, Application US/08961527  
Publication No. US20020032323A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 205:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2395 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-205

Query Match 86.7%; Score 15.6; DB 8; Length 2395;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAGTSCCGGCGCATG 18  
|||||:|||||  
DB 1257 GACCAGTCCCGGCGCATG 1240

RESULT 17  
US-10-158-844-205/c  
Sequence 205, Application US/10158844  
Publication No. US20040029118A1  
GENERAL INFORMATION:  
APPLICANT: Kunsch et al.  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville

```
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-R
COMPUTER: Dell Latitude Pentium 3
OPERATING SYSTEM: Windows 98
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/158.844
  FILING DATE: 03-Jun-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/961,527
  FILING DATE: 1997-10-30
  APPLICATION NUMBER: US 60/029,960
  FILING DATE: 1996-10-31
ATTORNEY/AGENT INFORMATION:
  NAME: Hyman, Mark J.
  REGISTRATION NUMBER: 46,789
  REFERENCE/DOCKET NUMBER: PB340P1D1
INFORMATION FOR SEQ ID NO: 205:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 2395 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
  SEQUENCE DESCRIPTION: SEQ ID NO: 205:
US-10-158-844-205

Query Match      86.7%; Score 15.6; DB 17; Length 2395;
Best Local Similarity 83.3%; Pred. No. 2,1e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 1 GACCACTSCCGGCATG 18
Db 1257 GACCACGCGCCGACATG 1240

RESULT 18
US-10-027-632-174581
; Sequence 174581, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174581
; LENGTH: 713059
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(713059)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174581
```

```
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174581

Query Match      86.7%; Score 15.6; DB 13; Length 713059;
Best Local Similarity 83.3%; Pred. No. 69;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 1 GACCACTSCCGGCATG 18
Db 462752 GACCACGCGCCGACATG 462769

RESULT 19
US-10-027-632-174581
; Sequence 174581, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174581
; LENGTH: 713059
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(713059)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174581

Query Match      86.7%; Score 15.6; DB 17; Length 713059;
Best Local Similarity 83.3%; Pred. No. 69;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 1 GACCACTSCCGGCATG 18
Db 462752 GACCACGCGCCGACATG 462769

RESULT 20
US-10-472-928-4979/c
; Sequence 4979, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W0
; CURRENT APPLICATION NUMBER: US/10/472,928
; PRIOR FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: Seqwin99, version 1.03
```



SEQ ID NO 4979  
LENGTH: 2162598  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
US-10-472-928-4979

Query Match 86.7%; Score 15.6; DB 21; Length 2162598;  
Best Local Similarity 83.3%; Pred. No. 55;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCCGGCATG 18  
Db 190126 GACCACTGSCCGGCATG 190109

RESULT 21  
US-10-156-761-6136/c  
Sequence 6136, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 6136  
LENGTH: 330  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(330)  
US-10-156-761-6136

Query Match 84.4%; Score 15.2; DB 15; Length 330;  
Best Local Similarity 87.5%; Pred. No. 4.9e+02;  
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGTSCCGGCATG 18  
Db 193 CCACGTSCCGGCATG 178

RESULT 22  
US-10-461-862-83/c  
Sequence 83, Application US/10461862  
Publication No. US2005090434A1  
GENERAL INFORMATION:  
APPLICANT: David W. Morris  
APPLICANT: Marc S. Malandro  
TITLE OF INVENTION: Novel Therapeutic Targets in Cancer  
FILE REFERENCE: 529452001800  
CURRENT APPLICATION NUMBER: US/10/461,862  
CURRENT FILING DATE: 2003-06-13  
NUMBER OF SEQ ID NOS: 184  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 83  
LENGTH: 110021  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc\_feature

LOCATION: (1)....(110021)  
OTHER INFORMATION: n = A,T,C or G  
US-10-461-862-83

Query Match 84.4%; Score 15.2; DB 21; Length 110021;  
Best Local Similarity 87.5%; Pred. No. 1.6e+02;  
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGTSCCGGCATG 18  
Db 41825 CCACGTSCCGGCATG 41810

RESULT 23  
US-09-294-093B-4374  
Sequence 4374, Application US/09294093B  
Patent No. US2001005135A1  
GENERAL INFORMATION:  
APPLICANT: Ito, Laura, Y.  
APPLICANT: Sherman, Bradley, K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL  
FILE REFERENCE: PL-0009 US  
CURRENT APPLICATION NUMBER: US/09/294,093B  
CURRENT FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: 60/082,567  
PRIOR FILING DATE: April 21, 1998  
NUMBER OF SEQ ID NOS: 6207  
SOFTWARE: PERL Program  
SEQ ID NO 4374  
LENGTH: 209  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. US2001005135A1 700353928H1  
LOCATION: 36, 63, 114, 134  
OTHER INFORMATION: a, t, c, g, or other  
US-09-294-093B-4374

Query Match 81.1%; Score 14.6; DB 9; Length 209;  
Best Local Similarity 82.4%; Pred. No. 1.1e+03;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCCGGCAT 17  
Db 87 GACCACTGSCCGGCAT 103

RESULT 24  
US-10-425-115-76979  
Sequence 76979, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 76979  
LENGTH: 271  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MT4577\_170231C.1  
US-10-425-115-76979



LOCATION: (478) .. (478)  
OTHER INFORMATION: n = any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (528) .. (528)  
OTHER INFORMATION: n = any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (539) .. (539)  
OTHER INFORMATION: n = any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (553) .. (553)  
OTHER INFORMATION: n = any nucleotide  
US-10-259-165-501

Query Match 81.1%; Score 14.6; DB 15; Length 568;  
Best Local Similarity 82.4%; Pred. NO. 8.8e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGCAT 17  
|||||:|||||  
DB 378 GACCACTGTCGCGCAT 394

RESULT 29  
US-10-437-963-73433  
Sequence 73433, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovacic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21 (53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 73433  
LENGTH: 735  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_73714C.1  
US-10-437-963-73433

Query Match 81.1%; Score 14.6; DB 19; Length 735;  
Best Local Similarity 82.4%; Pred. NO. 8.4e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCACGTSCGCGCATG 18  
|||||:|||||  
DB 317 ACCACGTSCGCGCATG 333

RESULT 30  
US-10-107-431-2  
Sequence 2, Application US/10107431  
Publication No. US20030224364A1  
GENERAL INFORMATION:  
APPLICANT: Farnet, Chris  
APPLICANT: Stafea, Alfredo  
APPLICANT: Zazopoulos, Emmanuel  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
TITLE OF INVENTION: BIOSYNTHETIC LOCI  
FILE REFERENCE: 3001-7US  
CURRENT APPLICATION NUMBER: US/10/107,431

CURRENT FILING DATE: 2002-03-28  
NUMBER OF SEQ ID NOS: 282  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
LENGTH: 1041  
TYPE: DNA  
ORGANISM: Streptomyces mobaraensis  
US-10-107-431-2

Query Match 81.1%; Score 14.6; DB 17; Length 1041;  
Best Local Similarity 82.4%; Pred. NO. 7.8e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGCAT 17  
|||||:|||||  
DB 697 GACCACTGTCGCGCAT 713

Search completed: July 20, 2005, 22:08:23  
Job time : 466.857 secs

**THIS PAGE IS BLANK**



101	14	77.8	1701	4	US-09-949-016-2634	Sequence 2634, Ap	174	13.6	75.6	601	4	US-09-949-016-34853	Sequence 34853, A
102	14	77.8	1879	4	US-09-023-655-742	Sequence 742, App	175	13.6	75.6	601	4	US-09-949-016-34854	Sequence 34854, A
103	14	77.8	1908	1	US-08-173-508-1	Sequence 1, Appl1	176	13.6	75.6	601	4	US-09-949-016-73464	Sequence 73464, A
104	14	77.8	1908	2	US-08-265-310-1	Sequence 1, Appl1	177	13.6	75.6	601	4	US-09-949-016-73465	Sequence 73465, A
105	14	77.8	1908	3	US-08-951-742-1	Sequence 1, Appl1	178	13.6	75.6	601	4	US-09-949-016-121671	Sequence 121671, A
106	14	77.8	1920	2	US-08-855-714-1	Sequence 1, Appl1	179	13.6	75.6	601	4	US-09-949-016-135152	Sequence 135152, A
107	14	77.8	1980	4	US-09-252-991A-4190	Sequence 4190, Ap	180	13.6	75.6	601	4	US-09-949-016-135153	Sequence 135153, A
108	14	77.8	1980	3	US-09-491-522-6	Sequence 6, Appl1	181	13.6	75.6	741	4	US-08-252-991A-4070	Sequence 4070, Ap
109	14	77.8	2034	4	US-09-902-540-8514	Sequence 8514, Ap	182	13.6	75.6	885	4	US-09-902-540-5476	Sequence 5476, Ap
110	14	77.8	2070	4	US-09-252-991A-2189	Sequence 2189, Ap	183	13.6	75.6	903	4	US-09-902-540-4758	Sequence 4758, Ap
111	14	77.8	2312	4	US-09-902-540-3277	Sequence 3277, Ap	184	13.6	75.6	1164	4	US-09-252-991A-13302	Sequence 13302, A
112	14	77.8	2316	4	US-09-902-540-3869	Sequence 3869, Ap	185	13.6	75.6	1173	4	US-09-252-991A-13384	Sequence 13384, A
113	14	77.8	2450	3	US-09-491-522-2	Sequence 2, Appl1	186	13.6	75.6	1182	4	US-09-489-039A-82	Sequence 82, Appl1
114	14	77.8	2540	1	US-08-027-986-4	Sequence 4, Appl1	187	13.6	75.6	1419	4	US-08-252-991A-13796	Sequence 13796, A
115	14	77.8	2582	2	US-08-816-105A-2	Sequence 2, Appl1	188	13.6	75.6	1419	4	US-09-902-540-8278	Sequence 8278, Ap
116	14	77.8	3636	4	US-09-949-016-5530	Sequence 5530, Ap	189	13.6	75.6	1469	4	US-09-949-016-19683	Sequence 4983, Ap
117	14	77.8	5179	4	US-09-674-866A-1	Sequence 1, Appl1	190	13.6	75.6	1470	4	US-09-270-767-10865	Sequence 10865, A
118	14	77.8	5504	4	US-09-902-540-716	Sequence 716, App	191	13.6	75.6	1569	4	US-09-252-991A-12675	Sequence 12675, A
119	14	77.8	5883	4	US-09-949-016-5001	Sequence 5001, Ap	192	13.6	75.6	1647	4	US-09-902-540-5342	Sequence 5342, Ap
120	14	77.8	6692	3	US-09-491-522-1	Sequence 1, Appl1	193	13.6	75.6	1766	4	US-08-252-991A-4139	Sequence 4129, Ap
121	14	77.8	6855	4	US-09-902-540-897	Sequence 897, App	194	13.6	75.6	1794	4	US-09-902-540-3791	Sequence 3791, Ap
122	14	77.8	7626	4	US-09-976-594-14	Sequence 14, Appl1	195	13.6	75.6	1854	4	US-09-270-767-11950	Sequence 11900, A
123	14	77.8	7626	4	US-09-919-039-14	Sequence 14, Appl1	196	13.6	75.6	1995	4	US-09-252-991A-13546	Sequence 13546, A
124	14	77.8	7626	4	US-09-902-540-1151	Sequence 1151, Ap	197	13.6	75.6	2119	3	US-09-032-372-6	Sequence 6, Appl1
125	14	77.8	18187	4	US-09-949-016-14126	Sequence 14126, A	198	13.6	75.6	2508	4	US-08-252-991A-793	Sequence 793, App
126	14	77.8	18187	4	US-09-949-016-14127	Sequence 14127, A	199	13.6	75.6	2577	4	US-09-252-991A-828	Sequence 828, App
127	14	77.8	18700	4	US-09-949-016-13140	Sequence 13140, A	200	13.6	75.6	2730	4	US-08-489-039A-6744	Sequence 6744, App
128	14	77.8	21143	4	US-09-902-540-1171	Sequence 1171, Ap	201	13.6	75.6	2871	4	US-09-949-016-13998	Sequence 3998, Ap
129	14	77.8	24905	4	US-09-902-540-1225	Sequence 1225, Ap	202	13.6	75.6	2934	4	US-09-252-991A-11690	Sequence 11690, A
130	14	77.8	24905	4	US-09-949-016-13288	Sequence 13288, Ap	203	13.6	75.6	3650	4	US-09-620-312D-30	Sequence 30, Appl1
131	14	77.8	24984	4	US-09-949-016-13288	Sequence 13288, A	204	13.6	75.6	3764	4	US-09-949-016-4511	Sequence 4511, Ap
132	14	77.8	24984	4	US-09-949-016-13289	Sequence 13289, A	205	13.6	75.6	5581	4	US-09-023-655-966	Sequence 966, App
133	14	77.8	28054	4	US-09-902-540-1248	Sequence 1248, Ap	206	13.6	75.6	6604	4	US-09-949-016-16785	Sequence 16725, A
134	14	77.8	36620	4	US-09-952-060-30	Sequence 30, Appl1	207	13.6	75.6	6763	4	US-09-949-016-13399	Sequence 3399, Ap
135	14	77.8	43280	2	US-08-804-227C-1	Sequence 1, Appl1	208	13.6	75.6	8030	4	US-09-902-540-857	Sequence 857, App
136	14	77.8	4689	1	US-08-471-119A-1	Sequence 1, Appl1	209	13.6	75.6	12023	4	US-09-949-016-16253	Sequence 16253, A
137	14	77.8	47981	4	US-09-679-279-1	Sequence 1, Appl1	210	13.6	75.6	12241	4	US-09-948-138-4	Sequence 4, Appl1
138	14	77.8	49301	4	US-09-949-016-16296	Sequence 16296, A	211	13.6	75.6	13737	3	US-09-538-414-10	Sequence 10, Appl1
139	14	77.8	70563	4	US-09-949-016-16743	Sequence 16743, A	212	13.6	75.6	13737	4	US-10-074-279-9	Sequence 10, Appl1
140	14	77.8	75395	4	US-09-984-890-3	Sequence 3, Appl1	213	13.6	75.6	13857	4	US-09-620-312D-75	Sequence 75, Appl1
141	14	77.8	75395	4	US-10-274-194-3	Sequence 3, Appl1	214	13.6	75.6	14446	4	US-09-810-861B-4	Sequence 4, Appl1
142	14	77.8	91933	4	US-09-949-016-11855	Sequence 11855, A	215	13.6	75.6	14705	4	US-09-949-016-16249	Sequence 16249, A
143	14	77.8	91933	4	US-09-949-016-14628	Sequence 14628, A	216	13.6	75.6	18551	4	US-09-902-540-1187	Sequence 1187, Ap
144	14	77.8	939353	4	US-09-949-016-14573	Sequence 14573, A	217	13.6	75.6	24333	4	US-09-639-207-9	Sequence 9, Appl1
145	14	77.8	939353	4	US-09-949-016-14574	Sequence 14574, A	218	13.6	75.6	26492	4	US-09-902-540-1234	Sequence 1234, Ap
146	14	77.8	818128	4	US-09-949-016-14546	Sequence 14546, A	219	13.6	75.6	29433	4	US-09-949-016-15740	Sequence 15740, A
147	14	77.8	818128	4	US-09-949-016-14547	Sequence 14547, A	220	13.6	75.6	29899	4	US-09-902-540-1265	Sequence 1265, Ap
148	14	77.8	818128	4	US-09-949-016-14548	Sequence 14548, A	221	13.6	75.6	34534	4	US-09-949-016-15141	Sequence 15141, A
149	14	77.8	818128	4	US-09-949-016-14549	Sequence 14549, A	222	13.6	75.6	35399	4	US-09-902-540-1260	Sequence 1260, Ap
150	14	77.8	818128	4	US-09-949-016-14550	Sequence 14550, A	223	13.6	75.6	36457	4	US-09-949-016-12404	Sequence 13404, A
151	14	77.8	818128	4	US-09-949-016-14551	Sequence 14551, A	224	13.6	75.6	36457	4	US-09-949-016-13031	Sequence 13031, A
152	14	77.8	818128	4	US-09-949-016-14552	Sequence 14552, A	225	13.6	75.6	36620	4	US-09-952-060-31	Sequence 31, Appl1
153	14	77.8	818128	4	US-09-949-016-14553	Sequence 14553, A	226	13.6	75.6	36778	3	US-09-105-537-5	Sequence 5, Appl1
154	14	77.8	818128	4	US-09-949-016-14554	Sequence 14554, A	227	13.6	75.6	37215	4	US-09-949-016-15526	Sequence 15526, A
155	14	77.8	818128	4	US-09-949-016-14555	Sequence 14555, A	228	13.6	75.6	37474	4	US-09-952-060-26	Sequence 26, Appl1
156	14	77.8	818128	4	US-09-949-016-14556	Sequence 14556, A	229	13.6	75.6	38506	3	US-09-320-878-19	Sequence 19, Appl1
157	14	77.8	818128	4	US-09-949-016-14557	Sequence 14557, A	230	13.6	75.6	38506	4	US-09-141-908-1	Sequence 1, Appl1
158	14	77.8	818128	4	US-09-949-016-14558	Sequence 14558, A	231	13.6	75.6	38506	4	US-09-657-440-19	Sequence 19, Appl1
159	14	77.8	818128	4	US-09-949-016-14559	Sequence 14559, A	232	13.6	75.6	38519	4	US-09-952-060-29	Sequence 29, Appl1
160	14	77.8	818128	4	US-09-949-016-14560	Sequence 14560, A	233	13.6	75.6	88617	4	US-09-949-016-12524	Sequence 12524, A
161	14	77.8	818128	4	US-09-949-016-14561	Sequence 14561, A	234	13.6	75.6	99439	4	US-09-949-016-13557	Sequence 13597, A
162	14	77.8	818128	4	US-09-949-016-14562	Sequence 14562, A	235	13.6	75.6	134708	4	US-09-949-016-13577	Sequence 13841, A
163	14	77.8	818128	4	US-09-949-016-14564	Sequence 14564, A	236	13.6	75.6	154746	4	US-09-827-688-8	Sequence 8, Appl1
164	14	77.8	818128	4	US-09-949-016-14565	Sequence 14565, A	237	13.6	75.6	154746	4	US-09-827-688-8	Sequence 8, Appl1
165	14	77.8	818128	4	US-09-949-016-14566	Sequence 14566, A	238	13.6	75.6	174029	4	US-09-949-016-12610	Sequence 12610, A
166	14	77.8	818128	4	US-09-949-016-14567	Sequence 14567, A	239	13.6	75.6	174030	4	US-09-949-016-13880	Sequence 13880, A
167	14	77.8	4403765	3	US-09-103-840A-2	Sequence 2, Appl1	240	13.6	75.6	223354	4	US-09-705-400-64	Sequence 64, Appl1
168	14	77.8	4411529	3	US-09-103-840A-1	Sequence 1, Appl1	241	13.6	75.6	238815	4	US-09-949-016-16274	Sequence 16274, A
169	13.6	75.6	289	4	US-09-270-767-12817	Sequence 12817, A	242	13.6	75.6	300402	4	US-09-949-016-13632	Sequence 13632, A
170	13.6	75.6	299	4	US-09-313-294A-6210	Sequence 6210, Ap	243	13.6	75.6	784019	4	US-09-949-016-14033	Sequence 14033, A
171	13.6	75.6	376	4	US-09-621-976-8013	Sequence 8013, Ap	244	13.6	75.6	826152	4	US-09-949-016-12777	Sequence 12777, A
172	13.6	75.6	414	4	US-09-252-991A-4060	Sequence 4060, Ap	245	13.4	74.4	438	4	US-09-252-991A-15524	Sequence 15524, A
173	13.6	75.6	564	4	US-09-252-991A-4096	Sequence 4096, Ap	246	13.4	74.4	601	4	US-09-949-016-29378	Sequence 29378, A

C 247	13.4	74.4	601	4	US-09-949-016-29379	Sequence 29379, A
C 248	13.4	74.4	601	4	US-09-949-016-35261	Sequence 35261, A
C 249	13.4	74.4	601	4	US-09-949-016-35262	Sequence 35262, A
C 250	13.4	74.4	601	4	US-09-949-016-154544	Sequence 154544, A
C 251	13.4	74.4	601	4	US-09-949-016-154545	Sequence 154545, A
C 252	13.4	74.4	601	4	US-09-949-016-154546	Sequence 154546, A
C 253	13.4	74.4	601	4	US-09-949-016-154547	Sequence 154547, A
C 254	13.4	74.4	601	4	US-09-949-016-156823	Sequence 156823, A
C 255	13.4	74.4	601	4	US-09-949-016-156824	Sequence 156824, A
C 256	13.4	74.4	601	4	US-09-949-016-173843	Sequence 173843, A
C 257	13.4	74.4	601	4	US-09-949-016-173844	Sequence 173844, A
C 258	13.4	74.4	972	4	US-09-252-991A-6718	Sequence 6718, A
C 259	13.4	74.4	1212	4	US-09-252-991A-6685	Sequence 6685, A
C 260	13.4	74.4	1230	4	US-09-902-540-5591	Sequence 5591, A
C 261	13.4	74.4	1335	4	US-09-949-016-2469	Sequence 2469, A
C 262	13.4	74.4	1380	4	US-09-252-991A-6613	Sequence 6613, A
C 263	13.4	74.4	1380	4	US-09-252-991A-15862	Sequence 15862, A
C 264	13.4	74.4	1836	4	US-09-252-991A-15831	Sequence 15831, A
C 265	13.4	74.4	2055	4	US-09-252-991A-15771	Sequence 15771, A
C 266	13.4	74.4	2106	4	US-09-252-991A-15891	Sequence 15891, A
C 267	13.4	74.4	7070	4	US-09-949-016-12469	Sequence 12469, A
C 268	13.4	74.4	7070	4	US-09-949-016-15322	Sequence 15322, A
C 269	13.4	74.4	9306	3	US-09-453-7028-231	Sequence 231, A
C 270	13.4	74.4	11879	4	US-09-949-016-14211	Sequence 14211, A
C 271	13.4	74.4	14861	4	US-09-902-540-1127	Sequence 1127, A
C 272	13.4	74.4	15222	4	US-09-949-016-11916	Sequence 11916, A
C 273	13.4	74.4	15223	4	US-09-949-016-16912	Sequence 16912, A
C 274	13.4	74.4	18319	4	US-09-949-016-17446	Sequence 17446, A
C 275	13.4	74.4	26423	4	US-09-949-016-15459	Sequence 15459, A
C 276	13.4	74.4	38653	4	US-09-922-445-1	Sequence 1, A
C 277	13.4	74.4	40576	4	US-09-949-016-12329	Sequence 12329, A
C 278	13.4	74.4	40577	4	US-09-949-016-16663	Sequence 16663, A
C 279	13.4	74.4	41927	4	US-09-902-540-1268	Sequence 1268, A
C 280	13.4	74.4	50797	4	US-09-949-016-16346	Sequence 16346, A
C 281	13.4	74.4	50797	4	US-09-949-016-16347	Sequence 16347, A
C 282	13.4	74.4	56448	4	US-09-949-016-16463	Sequence 16463, A
C 283	13.4	74.4	66627	4	US-09-949-016-12630	Sequence 12630, A
C 284	13.4	74.4	66628	4	US-09-949-016-16112	Sequence 16112, A
C 285	13.4	74.4	125672	4	US-09-949-016-16956	Sequence 16956, A
C 286	13.4	74.4	251672	4	US-09-949-016-17236	Sequence 17236, A
C 287	13.4	74.4	275110	4	US-09-949-016-11973	Sequence 11973, A
C 288	13.4	74.4	275110	4	US-09-949-016-12706	Sequence 12706, A
C 289	13.4	74.4	422592	4	US-09-949-016-16070	Sequence 16070, A
C 290	13.4	74.4	422592	4	US-09-949-016-14182	Sequence 14182, A
C 291	13.2	73.3	23	3	US-08-155-005A-15	Sequence 15, A
C 292	13.2	73.3	23	3	US-08-363-783-15	Sequence 15, A
C 293	13.2	73.3	23	4	US-08-661-758A-15	Sequence 15, A
C 294	13.2	73.3	23	4	US-09-304-232-151	Sequence 151, A
C 295	13.2	73.3	239	4	US-09-016-434-463	Sequence 463, A
C 296	13.2	73.3	345	4	US-09-252-991A-15649	Sequence 15649, A
C 297	13.2	73.3	388	4	US-09-513-999C-8528	Sequence 8528, A
C 298	13.2	73.3	433	4	US-09-621-976-12386	Sequence 12386, A
C 299	13.2	73.3	501	1	US-08-318-193-76	Sequence 76, A
C 300	13.2	73.3	601	4	US-09-949-016-78610	Sequence 78610, A

## ALIGNMENTS

RESULT 1  
US-09-199-637A-155  
Sequence 155, Application US/09199637A  
Patent No. 6355411  
GENERAL INFORMATION:  
APPLICANT: Ausubel, Frederick  
APPLICANT: Goodman, Howard M.  
APPLICANT: Rahme, Laurence G.  
APPLICANT: Mahajan-Miklos, Shalina  
APPLICANT: Cao, Man-Wah  
APPLICANT: Geo, Hui  
APPLICANT: Drenkard, Eliana  
APPLICANT: Tsongalis, John  
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID

TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
FILE REFERENCE: 00786/361002  
CURRENT APPLICATION NUMBER: US/09199, 637A  
CURRENT FILING DATE: 1998-11-25  
PRIOR APPLICATION NUMBER: 60/066, 517  
PRIOR FILING DATE: 1997-11-25  
NUMBER OF SEQ ID NOS: 437  
SOFTWARE: PASTSEQ for Windows Version 4.0  
SEQ ID NO 155  
LENGTH: 513  
TYPE: DNA  
ORGANISM: *Pseudomonas aeruginosa*  
US-09-199-637A-155

Query Match 86.7%; Score 15.6; DB 3; Length 513;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 15; Mismatches 2; Indels 0; Gaps 0;

Db 133 GACCACGTCCGCGCATG 150

OY 1 GACCACGTCCGCGCATG 18  
|||||||:|||||

RESULT 2  
US-09-107-433-417/C  
Sequence 417, Application US/09107433  
Patent No. 6800744  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS  
THERAPEUTICS

NUMBER OF SEQUENCES: 5206  
CORRESPONDENCE ADDRESS:  
ADDRESSER: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: <Unknown>  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09107, 433  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/ 085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 417:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 612 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: *Streptococcus pneumoniae*  
FEATURE:  
NAME/KEY: misc\_feature





US-09-180-109A-5

Sequence 5, Application US/09180109A  
Patent No. 6410293  
GENERAL INFORMATION:  
APPLICANT: MUKIMOTO, Fujio  
APPLICANT: NISHIO, Shoichi  
APPLICANT: AKIMARU, Jiro  
APPLICANT: MITSUDA, Satechi  
TITLE OF INVENTION: DNA Fragments Containi  
TITLE OF INVENTION: Use of the Same  
FILE REFERENCE: 0152-0490P  
CURRENT APPLICATION NUMBER: US/09/180,109A  
CURRENT FILING DATE: 1998-12-03  
PRIOR APPLICATION NUMBER: 09/047838 JAPAN  
PRIOR FILING DATE: 1997-03-03  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 1408  
TYPE: DNA  
ORGANISM: *Sphingomonas* sp.  
FEATURE:  
OTHER INFORMATION: Strain = SC42405  
NAME/KEY: CDS  
LOCATION: (202)..(1362)  
US-09-180-109A-5

```
Query Match      86.7%; Score 15.6; DB 3; Length 1408;
Beet Local Similarity 83.3%; Pred. No. 3e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```

Qy      1 GACCACGTSCCGGCATG 18
        |||||:|||||
Db      1057 GACCTCGTCCGGGCATG 1074

```

```

RESULT 8
US-09-180-109A-7
Sequence 7, Application US/09180109A
Patent No. 6410293
GENERAL INFORMATION:
APPLICANT: MUKUMOTO, Fujio
APPLICANT: NISHIO, Shoichi
APPLICANT: AKIMARU, Tiro
APPLICANT: MITSUDA, Satoshi
TITLE OF INVENTION: DNA Fragments Containin
TITLE OF INVENTION: Use of the Same
FILE REFERENCE: 0152-0490P
CURRENT APPLICATION NUMBER: US/09/180,109A
CURRENT FILING DATE: 1998-12-03
PRIOR APPLICATION NUMBER: 09/047838 JAPAN
PRIOR FILING DATE: 1997-03-03
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1408
TYPE: DNA
ORGANISM: Sphingomonas sp.
FEATURES:
OTHER INFORMATION: Strain = SC42405
NAME/KEY: CDS
LOCATION: (202)..(1362)
US-09-180-109A-7

```

	Query Match	Similarity	Score	DB	Length
Best Local	86.7%	83.3%	15.6	3	1408
Matches	15	Conservative	2	Mismatches	1
				Indels	0
				Gaps	0

```

QY      1 GACCACGTSCCSGGCATG 18
        |||||:|:|||||
Db      1057 GACCTCGTCCGGGCATG 1074

```

RESULT 9  
US-09-199-637A-136

```

Sequence 136, Application US/09199637A
Patent No. 6355411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahme, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
APPLICANT: Tan, Man-Wan
APPLICANT: Cao, Hui
APPLICANT: Drenkard, Eliana
APPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
FILE REFERENCE: 00786/361002
CURRENT APPLICATION NUMBER: US/09/199,637A
CURRENT FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: 60/066,517
PRIOR FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 136
LENGTH: 2048
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-136

```

Query Match	86.7%;	Score 15.6;	DB 3;	Length 2048;
Best Local Similarity	83.3%;	Pred. No. 2.9e+02;		
Matches 15; Conservative	2;	Mismatches 1;	Indels 0;	Gaps 0;

```

QY      1 GACCACGTSCCGGCATG 18
        |||||:|:|||||
Db      1069 GACCACGTCCGGCATG 1086

```

RESULT 10  
 US-08-961-527--205/c  
 Sequence 205, Application US/08961527  
 Patent No. 6420135  
 GENERAL INFORMATION:  
 APPLICANT: Charles Kunisch  
 TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 NUMBER OF SEQUENCES: 391  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 inch, 1.4mb storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/961,527  
 FILING DATE:  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Brookes, A. Anders  
 REGISTRATION NUMBER: 36,373  
 REFERENCE/DOCKET NUMBER: PB340P1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 309-8504  
 TELEFAX: (301) 309-8512  
 INFORMATION FOR SEQ ID NO: 205:  
 SEQUENCE CHARACTERISTICS:

LENGTH: 2395 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-205

Query Match 86.7%; Score 15.6; DB 3; Length 2395;  
Best Local Similarity 83.3%; Pred. No. 2.9e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1257 GACCACGTGCCGCATG 1240

RESULT 11  
US-09-949-016-1079  
Sequence 1079, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1079  
LENGTH: 4927  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-1079

Query Match 86.7%; Score 15.6; DB 4; Length 4927;  
Best Local Similarity 83.3%; Pred. No. 2.7e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 24 GACCACGGCCGGGCATG 41

RESULT 12  
US-09-949-016-15371/C  
Sequence 15371, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15371  
LENGTH: 304533  
TYPE: DNA  
ORGANISM: Human  
FEATURE:

NAME/KEY: misc\_feature  
LOCATION: (1)...(304533)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15371

Query Match 86.7%; Score 15.6; DB 4; Length 304533;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 299865 GACCACGTGCCGCATG 299848

RESULT 13  
US-09-949-016-15372/C  
Sequence 15372, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15372  
LENGTH: 304533  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(304533)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15372

Query Match 86.7%; Score 15.6; DB 4; Length 304533;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 299865 GACCACGTGCCGCATG 299848

RESULT 14  
US-09-949-016-17330  
Sequence 17330, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17330  
LENGTH: 10948

TYPE: DNA  
ORGANISM: Human  
US-09-949-016-17310

Query Match 83.3%; Score 15; DB 4; Length 10946;  
Best Local Similarity 88.2%; Pred. No. 5e+02;  
Matches 15; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCACGTSCCGGCATG 18  
|||||:|||||  
Db 4160 ACCACGTCCCTCGCATG 4176

RESULT 15  
US-09-621-976-16915/c  
Sequence 16915, Application US/09621976  
Patent No. 6639063  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Jobert, S.  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: GENSET 054PR2  
CURRENT APPLICATION NUMBER: US/09/621,976  
CURRENT FILING DATE: 2000-07-21  
NUMBER OF SEQ ID NOS: 19335  
SOFTWARE: Patent.pm  
SEQ ID NO 16915  
LENGTH: 510  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-621-976-16915

Query Match 81.1%; Score 14.6; DB 4; Length 510;  
Best Local Similarity 82.4%; Pred. No. 9.1e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACGTSCCGGCAT 17  
|||||:|||||  
Db 59 GACCACGTGCCGTCAT 43

RESULT 16  
US-09-640-211A-238/c  
Sequence 238, Application US/09640211A  
Patent No. 6833446  
GENERAL INFORMATION:  
APPLICANT: Wood, Marion  
APPLICANT: Shenk, Michael A.  
APPLICANT: McGrath, Annette  
APPLICANT: Glenn, Matthew  
TITLE OF INVENTION: Compositions and Methods for the  
FILE REFERENCE: 11000.1021CIU  
CURRENT APPLICATION NUMBER: US/09/640,211A  
CURRENT FILING DATE: 2000-08-16  
NUMBER OF SEQ ID NOS: 2368  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 238  
LENGTH: 521  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-640-211A-238

Query Match 81.1%; Score 14.6; DB 4; Length 521;  
Best Local Similarity 82.4%; Pred. No. 9.1e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACGTSCCGGCAT 17  
|||||:|||||  
Db 497 GACCACGTCCCGGCGT 481

RESULT 17  
US-09-640-211A-1210/c  
Sequence 1210, Application US/09640211A  
Patent No. 6833446  
GENERAL INFORMATION:  
APPLICANT: Wood, Marion  
APPLICANT: Shenk, Michael A.  
APPLICANT: McGrath, Annette  
APPLICANT: Glenn, Matthew  
TITLE OF INVENTION: Compositions and Methods for the  
FILE REFERENCE: 11000.1021CIU  
CURRENT APPLICATION NUMBER: US/09/640,211A  
CURRENT FILING DATE: 2000-08-16  
NUMBER OF SEQ ID NOS: 2368  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1210  
LENGTH: 521  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-640-211A-1210

Query Match 81.1%; Score 14.6; DB 4; Length 521;  
Best Local Similarity 82.4%; Pred. No. 9.1e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACGTSCCGGCAT 17  
|||||:|||||  
Db 497 GACCACGTCCCGGCGT 481

RESULT 18  
US-09-902-540-7926  
Sequence 7926, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 7926  
LENGTH: 669  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-7926

Query Match 81.1%; Score 14.6; DB 4; Length 669;  
Best Local Similarity 82.4%; Pred. No. 8.9e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCACGTSCCGGCATG 18  
|||||:|||||  
Db 497 ACCACGTACCGGCATG 513

RESULT 19  
US-08-831-132-13/c  
Sequence 13, Application US/08831132  
Patent No. 6008322  
GENERAL INFORMATION:  
APPLICANT: Kuestner, Rolf B.  
APPLICANT: Conklin, Darrell C.  
APPLICANT: Lok, Si  
APPLICANT: Buddle, Michele  
APPLICANT: Downey, William  
TITLE OF INVENTION: STANNICALCIN-2

```

;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/831,132
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A.
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 96-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6672
;
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 888 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 1..72
;
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 73..888
;
; NAME/KEY: CDS
; LOCATION: 1..888
;
; US-08-831-132-13
;
; Query Match      81.1%; Score 14.6; DB 3; Length 888;
; Best Local Similarity 82.4%; Pred. No. 8.8e+02;
; Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; QY      1 GACCACGTSCCGGCAT 17
;         |||||:|:|:|
; DB      192 GCCCAGTCCCGGCAT 176
;
; RESULT 20
; US-09-416-150-13/c
; Sequence 13, Application US/09416150
; Patent No. 6171822
;
; GENERAL INFORMATION:
; APPLICANT: Kuestner, Rolf E.
;              Konklin, Darrell C.
;              Lok, Si
;              Buddle, Michele
;              Downey, William
;
; TITLE OF INVENTION: STANNIOCALCIN-2
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/416,150
; FILING DATE: 11-Oct-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/831,132
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A.
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 96-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6672
;
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 888 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 1..72
;
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 73..888
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..888
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
;
; US-09-416-150-13
;
; Query Match      81.1%; Score 14.6; DB 3; Length 888;
; Best Local Similarity 82.4%; Pred. No. 8.8e+02;
; Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; QY      1 GACCACGTSCCGGCAT 17
;         |||||:|:|:|
; DB      192 GCCCAGTCCCGGCAT 176
;
; RESULT 21
; US-09-902-540-7901/c
; Sequence 7901, Application US/09902540
; Patent No. 6833447
;
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
;              Hinkle, Gregory J.
;              Slater, Steven C.
;              Wiegand, Roger C.
;
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(11549)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 7901
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
;
; US-09-902-540-7901
;
; Query Match      81.1%; Score 14.6; DB 4; Length 1173;
; Best Local Similarity 82.4%; Pred. No. 8.6e+02;
; Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; QY      1 GACCACGTSCCGGCAT 17
;         |||||:|:|:|
; DB      684 GACCACGTCCCGGCCT 668

```

```
RESULT 22
US-09-902-540-7980/c
; Sequence 7980, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 7980
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-7980

Query Match      81.1%; Score 14.6; DB 4; Length 1347;
Best Local Similarity 82.4%; Pred. No. 8.6e+02;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACGTSCCGGCAT 17
Db      1149 GACCACGTGCGCGGCAT 1133

RESULT 23
US-09-270-767-15069
; Sequence 15069, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15069
; LENGTH: 1424
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15069

Query Match      81.1%; Score 14.6; DB 4; Length 1424;
Best Local Similarity 82.4%; Pred. No. 8.6e+02;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACGTSCCGGCAT 17
Db      751 GTCCACGTGCGCGGCAT 767

RESULT 24
US-09-821-167-13
; Sequence 13, Application US/09821167
; Patent No. 656968
; GENERAL INFORMATION:
; APPLICANT: Hosted Jr., Thomas J.
; APPLICANT: Hosan, Ann C.
; TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria
; FILE REFERENCE: INO11490
; CURRENT APPLICATION NUMBER: US/09/821,167
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/194,461
```

```
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1749
; TYPE: DNA
; ORGANISM: Micromonospora rosaria
US-09-821-167-13

Query Match      81.1%; Score 14.6; DB 4; Length 1749;
Best Local Similarity 82.4%; Pred. No. 8.5e+02;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACGTSCCGGCAT 17
Db      436 GACCACGTGCGCGGCAT 452

RESULT 25
US-09-949-016-2932
; Sequence 2932, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2932
; LENGTH: 3966
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2932

Query Match      81.1%; Score 14.6; DB 4; Length 3966;
Best Local Similarity 82.4%; Pred. No. 8.1e+02;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 ACCACGTSCCGGCATG 18
Db      2385 ACCACGTGCGCGGCATG 2401

RESULT 26
US-09-949-016-238
; Sequence 238, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 238
; LENGTH: 4979
```

;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-238

Query Match 81.1%; Score 14.6; DB 4; Length 4979;  
Best Local Similarity 82.4%; Pred. No. 8e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 ACCACGTSCCGGCATG 18  
|||||:|:|  
DB 4484 ACCACGTCCCCGCATG 4500

RESULT 27  
US-09-902-540-808  
; Sequence 808, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 808  
; LENGTH: 6012  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-808

Query Match 81.1%; Score 14.6; DB 4; Length 6012;  
Best Local Similarity 82.4%; Pred. No. 7.9e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACTSCCGGCAT 17  
|||||:|:|  
DB 1922 GACCACTCGCGGCAT 1938

RESULT 28  
US-09-902-540-799/c  
; Sequence 799, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 799  
; LENGTH: 6488  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-799

Query Match 81.1%; Score 14.6; DB 4; Length 6488;  
Best Local Similarity 82.4%; Pred. No. 7.8e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 ACCACGTSCCGGCATG 18  
|||||:|:|  
DB 3302 ACCACGTACCGGCATG 3286

RESULT 29  
US-09-902-540-796  
; Sequence 796, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 796  
; LENGTH: 6603  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-796

Query Match 81.1%; Score 14.6; DB 4; Length 6603;  
Best Local Similarity 82.4%; Pred. No. 7.8e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACTSCCGGCAT 17  
|||||:|:|  
DB 5510 GACCACTCCCCGCCT 5526

RESULT 30  
US-09-028-934-28  
; Sequence 28, Application US/09028934  
; Patent No. 6117670  
; GENERAL INFORMATION:  
; APPLICANT: Ligon, James M.  
; APPLICANT: Hill, Dwight S.  
; APPLICANT: Lam, Steven T.  
; APPLICANT: Hammer, Philip E.  
; APPLICANT: van Bee, Karl-Heinz  
; APPLICANT: Kirner, Sabine  
; APPLICANT: Young, Thomas R.  
; TITLE OF INVENTION: Pyrolytic Biosynthesis Genes and Uses  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: No. 6117670artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: NC  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/028,934  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/729,214  
; FILING DATE: 09-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/258,261  
; FILING DATE: 08-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: CGC1506/CIP7  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8931 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Burkholderia cepacia  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 657..2267  
OTHER INFORMATION: /product= "PrnA"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2270..3355  
OTHER INFORMATION: /product= "PrnB"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3421..5121  
OTHER INFORMATION: /product= "PrnC"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5145..6266  
OTHER INFORMATION: /product= "PrnD"  
US-09-028-934-28

Query Match 81.1%; Score 14.6; DB 3; Length 8931;  
Best Local Similarity 82.4%; Pred. No. 7.7e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACGACGTSCCGGCAT 17  
|||:|||||  
Db 2387 GACGACGTSCCGGCAT 2403

Search completed: July 20, 2005, 16:48:19  
Job time : 115.286 secs

**THIS PAGE IS BLANK**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 15:46:36 ; Search time 2237.57 Seconds  
(without alignments)  
306.206 Million cell updates/sec

Title: US-10-017-471b-1

Perfect score: 18

Sequence: 1 gaccacgtaccggcagc 18

Scoring table: IDENTITY NUC  
Gapop 10'-0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 300 summaries

Database :

EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_est3:  
4: gb\_est4:  
5: gb\_est5:  
6: gb\_est6:  
7: gb\_est7:  
8: gb\_est8:  
9: gb\_est9:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.2	95.6	742	9	CR810834
2	16.2	90.0	376	4	BI634046
3	16.2	90.0	703	6	CB016461
4	16.2	90.0	738	7	CN225729
5	15.6	86.7	223	2	BF813258
6	15.6	86.7	293	2	BB397065
7	15.6	86.7	295	6	CB995647
8	15.6	86.7	433	2	BF704768
9	15.6	86.7	435	2	BF656537
10	15.6	86.7	451	2	BB729006
11	15.6	86.7	510	9	CL797916
12	15.6	86.7	514	8	AQ397255
13	15.6	86.7	523	7	CO344004
14	15.6	86.7	539	1	AL903319
15	15.6	86.7	564	9	LBAPF3C09
16	15.6	86.7	581	5	BP221909
17	15.6	86.7	582	5	BP356698
18	15.6	86.7	583	5	BP280394
19	15.6	86.7	583	5	BP359091
20	15.6	86.7	587	5	BP364824
21	15.6	86.7	588	5	BP310516
22	15.6	86.7	588	8	AZ411523
23	15.6	86.7	602	1	AI389666
24	15.6	86.7	602	1	AI389666

25	15.6	86.7	612	6	CD311022
26	15.6	86.7	624	7	CN720685
27	15.6	86.7	631	6	CA100015
28	15.6	86.7	634	6	CA071958
29	15.6	86.7	636	7	CF367459
30	15.6	86.7	641	6	CB096877
31	15.6	86.7	655	6	CD629831
32	15.6	86.7	664	9	CL786031
33	15.6	86.7	666	6	CD629825
34	15.6	86.7	678	9	AG130601
35	15.6	86.7	701	6	CD629828
36	15.6	86.7	702	6	CD629833
37	15.6	86.7	713	6	CD629839
38	15.6	86.7	715	6	CD629829
39	15.6	86.7	727	9	CL720198
40	15.6	86.7	744	6	CD629841
41	15.6	86.7	744	9	CG675031
42	15.6	86.7	786	8	AQ689896
43	15.6	86.7	795	6	CB997742
44	15.6	86.7	799	5	BX616161
45	15.6	86.7	804	2	AW940798
46	15.6	86.7	853	8	B2553300
47	15.6	86.7	992	5	B0719203
48	15.6	86.7	999	5	B0188259
49	15.6	86.7	1000	9	CNS01086
50	15.6	86.7	1030	5	BM903853
51	15.6	86.7	1351	3	CR724745
52	15.6	86.7	1557	3	CR728715
53	15.6	86.7	1560	3	CR726317
54	15.6	86.7	1601	3	CR702207
55	15.6	86.7	1608	3	CR727784
56	15.6	86.7	1622	3	CR730580
57	15.6	86.7	1634	3	CR723875
58	15.6	86.7	1655	3	CR733878
59	15.2	84.4	359	8	AZ468435
60	15.2	84.4	467	8	AZ496471
61	15.2	84.4	552	2	BF077253
62	15.2	84.4	561	2	BE236993
63	15.2	84.4	566	2	BE237830
64	15.2	84.4	611	2	BF490349
65	15.2	84.4	686	9	CR027543
66	15.2	84.4	709	8	BH950057
67	15.2	84.4	714	4	BJ735213
68	15.2	84.4	719	4	BJ745844
69	15.2	84.4	738	4	BJ745844
70	15.2	84.4	777	8	BZ065629
71	15.2	84.4	819	8	BZ220159
72	15.2	84.4	849	9	CG100102
73	15.2	84.4	917	9	CG100099
74	15.2	84.4	1195	5	EM907019
75	15.2	84.4	184	4	BM484770
76	15.2	84.4	246	8	AZ888642
77	15.2	84.4	271	6	CA779153
78	15.2	84.4	297	7	CN029046
79	15.2	84.4	317	4	BI344931
80	15.2	84.4	317	4	BI345070
81	15.2	84.4	339	4	BI118321
82	15.2	84.4	395	6	CB547063
83	15.2	84.4	409	6	CB547063
84	15.2	84.4	412	1	AA609648
85	15.2	84.4	482	7	CF789783
86	15.2	84.4	496	5	EX926321
87	15.2	84.4	500	5	EX915300
88	15.2	84.4	546	9	CE122049
89	15.2	84.4	553	9	CE122049
90	15.2	84.4	554	6	CL866775
91	15.2	84.4	651	6	CL602170
92	15.2	84.4	672	2	BM539336
93	15.2	84.4	721	7	CK460965
94	15.2	84.4	791	5	BD1108056
95	15.2	84.4	791	6	BD1108056
96	15.2	84.4	837	9	CC494591
97	15.2	84.4	883	7	CO165983

C 98	15	83.3	926	6	CD791885	CD791885	EST63246	C 171	14.6	81.1	377	5	BY011149
C 99	15	83.3	936	6	CD779314	CD779314	EST650675	C 172	14.6	81.1	377	5	BY0117349
C 100	15	83.3	991	6	CNS024EB	AL180668	Tetracodon	C 173	14.6	81.1	377	5	BY205686
C 101	15	82.3	3612	3	AK053442	AK053442	Mus muscu	C 174	14.6	81.1	378	5	BY015000
C 102	14.8	82.2	914	1	AL574513	AL574513		C 175	14.6	81.1	378	5	BY016295
C 103	14.6	81.1	147	8	CC178278	CC178278	XE248 Bay	C 176	14.6	81.1	379	5	BY318960
C 104	14.6	81.1	170	7	CK098878	CK098878	A045P66.5	C 177	14.6	81.1	380	5	BY235292
C 105	14.6	81.1	178	5	BQ290364	BQ290364	hd79C05.9	C 178	14.6	81.1	381	5	BY229934
C 106	14.6	81.1	182	6	CA568095	CA568095	K0423F01-	C 179	14.6	81.1	381	7	CO946863
C 107	14.6	81.1	182	9	CE711187	CE711187	ligr-g8s-	C 180	14.6	81.1	383	5	BY013340
C 108	14.6	81.1	234	9	CES66467	CES66467	ligr-g8s-	C 181	14.6	81.1	383	5	BY068832
C 109	14.6	81.1	238	1	AA390339	AA390339	LD03283.5	C 182	14.6	81.1	383	5	BY175209
C 110	14.6	81.1	264	2	AM815496	AM815496	QV0-ST021	C 183	14.6	81.1	385	5	BY148921
C 111	14.6	81.1	280	8	B2477857	B2477857	BONON49TF	C 184	14.6	81.1	386	5	BY027567
C 112	14.6	81.1	278	8	AM454211	AM454211	zeH1101	C 185	14.6	81.1	387	2	BE433307
C 113	14.6	81.1	291	5	BY281797	BY281797		C 186	14.6	81.1	387	5	BY235317
C 114	14.6	81.1	294	2	AM428473	AM428473	66561 MAR	C 187	14.6	81.1	389	5	BY267409
C 115	14.6	81.1	295	6	CD553835	CD553835	B0366D12-	C 188	14.6	81.1	391	2	BB842330
C 116	14.6	81.1	298	5	BP138456	BP138456		C 189	14.6	81.1	393	5	BY278370
C 117	14.6	81.1	307	5	BY333719	BY333719		C 190	14.6	81.1	394	5	BY267085
C 118	14.6	81.1	310	5	BY129019	BY129019		C 191	14.6	81.1	395	7	CF944392
C 119	14.6	81.1	316	2	BE911620	BE911620	60163302	C 192	14.6	81.1	400	6	CB699634
C 120	14.6	81.1	327	6	BY782882	BY782882		C 193	14.6	81.1	400	6	CB699634
C 121	14.6	81.1	329	5	BY181592	BY181592		C 194	14.6	81.1	401	4	BY131409
C 122	14.6	81.1	329	5	BY191027	BY191027		C 195	14.6	81.1	404	5	BY131409
C 123	14.6	81.1	329	6	BY789463	BY789463		C 196	14.6	81.1	405	1	A1868523
C 124	14.6	81.1	332	5	BY325473	BY325473		C 197	14.6	81.1	408	5	BY019283
C 125	14.6	81.1	333	5	BY24375	BY24375	BY324375	C 198	14.6	81.1	409	5	BY019283
C 126	14.6	81.1	334	5	BY073543	BY073543		C 199	14.6	81.1	409	7	CF851819
C 127	14.6	81.1	334	5	BY29302	BY29302		C 200	14.6	81.1	410	7	N96719
C 128	14.6	81.1	340	5	BY189875	BY189875		C 201	14.6	81.1	414	6	CI19902
C 129	14.6	81.1	342	5	BY010190	BY010190		C 202	14.6	81.1	415	4	BM087299
C 130	14.6	81.1	342	5	BY213993	BY213993		C 203	14.6	81.1	417	7	BY273289
C 131	14.6	81.1	343	5	BY321572	BY321572		C 204	14.6	81.1	418	7	CM821667
C 132	14.6	81.1	344	2	BF511023	BF511023	UI-H-BT4-	C 205	14.6	81.1	419	2	CL213467
C 133	14.6	81.1	347	1	BY222040	BY222040		C 206	14.6	81.1	419	2	BB850083
C 134	14.6	81.1	348	1	AI089097	AI089097	CU81H12.8	C 207	14.6	81.1	419	4	BY568730
C 135	14.6	81.1	349	5	BY231642	BY231642		C 208	14.6	81.1	419	9	CI586760
C 136	14.6	81.1	350	5	BY044263	BY044263		C 209	14.6	81.1	420	5	BY275542
C 137	14.6	81.1	350	5	BY162806	BY162806		C 210	14.6	81.1	420	7	CM821472
C 138	14.6	81.1	350	5	BY355693	BY355693		C 211	14.6	81.1	423	5	BY223312
C 139	14.6	81.1	352	5	BY163418	BY163418		C 212	14.6	81.1	424	2	BB848780
C 140	14.6	81.1	353	5	BY043994	BY043994		C 213	14.6	81.1	424	5	BY281952
C 141	14.6	81.1	354	5	BY231635	BY231635		C 214	14.6	81.1	424	5	BY455103
C 142	14.6	81.1	355	1	AA735789	AA735789	CM09866.5	C 215	14.6	81.1	425	5	BY259533
C 143	14.6	81.1	355	1	BY324093	BY324093		C 216	14.6	81.1	426	2	BB850960
C 144	14.6	81.1	356	5	BY023197	BY023197		C 217	14.6	81.1	426	5	BY295370
C 145	14.6	81.1	356	5	BY044034	BY044034		C 218	14.6	81.1	428	6	CB794102
C 146	14.6	81.1	356	5	BY066574	BY066574		C 219	14.6	81.1	430	2	BB849148
C 147	14.6	81.1	356	5	BY182140	BY182140		C 220	14.6	81.1	431	5	BY241970
C 148	14.6	81.1	356	5	BY223601	BY223601		C 221	14.6	81.1	431	5	BY250665
C 149	14.6	81.1	358	5	BY071958	BY071958		C 222	14.6	81.1	431	7	CM819440
C 150	14.6	81.1	358	5	BY818198	BY818198		C 223	14.6	81.1	433	2	BB849589
C 151	14.6	81.1	358	5	BY203119	BY203119		C 224	14.6	81.1	433	5	BP087477
C 152	14.6	81.1	359	5	BY026243	BY026243		C 225	14.6	81.1	434	6	CA620032
C 153	14.6	81.1	359	5	BY231039	BY231039		C 226	14.6	81.1	435	2	AM358700
C 154	14.6	81.1	360	5	BY185924	BY185924		C 227	14.6	81.1	443	7	CF861873
C 155	14.6	81.1	360	5	BY187856	BY187856		C 228	14.6	81.1	443	7	CM819932
C 156	14.6	81.1	361	5	BY009542	BY009542		C 229	14.6	81.1	444	6	CB788315
C 157	14.6	81.1	361	5	BY013030	BY013030		C 230	14.6	81.1	444	8	CB793004
C 158	14.6	81.1	362	5	BY146507	BY146507		C 231	14.6	81.1	446	5	BY271817
C 159	14.6	81.1	362	5	BY176154	BY176154		C 232	14.6	81.1	446	9	EX907061
C 160	14.6	81.1	365	2	BF162208	BF162208	60177068	C 233	14.6	81.1	447	7	RI6064
C 161	14.6	81.1	365	5	BY088058	BY088058		C 234	14.6	81.1	452	7	CO284307
C 162	14.6	81.1	366	5	BY022227	BY022227		C 235	14.6	81.1	457	4	BC960052
C 163	14.6	81.1	368	5	BY235346	BY235346		C 236	14.6	81.1	457	6	CM542855
C 164	14.6	81.1	371	2	BB844425	BB844425		C 237	14.6	81.1	458	1	AA803236
C 165	14.6	81.1	371	5	BY012546	BY012546		C 238	14.6	81.1	458	7	CF896185
C 166	14.6	81.1	371	5	BY230476	BY230476		C 239	14.6	81.1	460	1	AA817646
C 167	14.6	81.1	372	5	BY014420	BY014420		C 240	14.6	81.1	467	2	BB857603
C 168	14.6	81.1	372	5	BY169527	BY169527		C 241	14.6	81.1	468	2	BB855178
C 169	14.6	81.1	374	2	BB842553	BB842553		C 242	14.6	81.1	470	2	BE751452
C 170	14.6	81.1	376	5	BY028119	BY028119		C 243	14.6	81.1	473	1	AA942179
CD791885								BY011149					
CD779314								BY0117349					
CNS024EB								BY205686					
AK053442								BY015000					
AL574513								BY318960					
CC178278								BY235292					
CK098878								BY229934					
BQ290364								CO946863					
CA568095								BY013340					
CE711187								BY068832					
CES66467								BY175209					
AA390339								BY148921					
AM815496								BY027567					
B2477857								BE433307					
AM454211								BY235317					
BY281797								BY267409					
AM428473								BB842330					
CD553835								BY278370					
BP138456								BY267085					
BY333719								CF944392					
BY129019								BY313265					
BE911620								CB699634					
BY782882								BT166039					
BY181592								BY131409					
BY191027								A1868523					
BY789463								BY264392					
BY325473								BY019283					
BY24375								CF851819					
BY073543								N96719					
BY29302								CI19902					
BY189875								BM087299					
BY010190								BY273289					
BY213993								CM821667					
BY321572								CL213467					
BF511023								BB850083					
BY222040								BY568730					
AI089097								CI586760					
BY231642								BY275542					
BY04													

Accession	Version	Source	Organism	Accession	Version	Source	Organism
C 244	14.6	81.1	473	1	AT1700719	we39g02.x	
C 245	14.6	81.1	473	2	BE021973	sm64g11.y	
C 246	14.6	81.1	474	2	BB863466	BB863466	
C 247	14.6	81.1	476	5	BX513312	BX513312	
C 248	14.6	81.1	479	2	BF765777	BF765777	
C 249	14.6	81.1	483	7	CO298673	CO298673	
C 250	14.6	81.1	486	4	CG633856	CG633856	
C 251	14.6	81.1	487	2	BF889536	BF889536	
C 252	14.6	81.1	488	6	CG726154	CG726154	
C 253	14.6	81.1	488	6	CG667743	CG667743	
C 254	14.6	81.1	490	7	CO719318	CO719318	
C 255	14.6	81.1	493	8	AZ853001	AZ853001	
C 256	14.6	81.1	500	9	CG540795	CG540795	
C 257	14.6	81.1	503	7	AA423593	AA423593	
C 258	14.6	81.1	507	1	AA78166	AA78166	
C 259	14.6	81.1	508	7	CG436786	CG436786	
C 260	14.6	81.1	509	9	CG720364	CG720364	
C 261	14.6	81.1	509	9	CG556189	CG556189	
C 262	14.6	81.1	510	8	BB860156	BB860156	
C 263	14.6	81.1	512	1	AA438361	AA438361	
C 264	14.6	81.1	514	6	CG548552	CG548552	
C 265	14.6	81.1	517	1	AI390058	AI390058	
C 266	14.6	81.1	517	4	BI340188	BI340188	
C 267	14.6	81.1	520	1	AV390930	AV390930	
C 268	14.6	81.1	521	1	BI641964	BI641964	
C 269	14.6	81.1	522	2	BB867564	BB867564	
C 270	14.6	81.1	522	3	CNS09676	CNS09676	
C 271	14.6	81.1	523	8	AZ583137	AZ583137	
C 272	14.6	81.1	524	7	CF578824	CF578824	
C 273	14.6	81.1	526	1	AI518223	AI518223	
C 274	14.6	81.1	529	7	BM176280	BM176280	
C 275	14.6	81.1	530	4	BI375435	BI375435	
C 276	14.6	81.1	530	4	CF862838	CF862838	
C 277	14.6	81.1	530	4	CF863617	CF863617	
C 278	14.6	81.1	538	7	CF862232	CF862232	
C 279	14.6	81.1	538	7	CF867985	CF867985	
C 280	14.6	81.1	540	1	AI531270	AI531270	
C 281	14.6	81.1	540	6	CA717718	CA717718	
C 282	14.6	81.1	542	2	AM213026	AM213026	
C 283	14.6	81.1	542	5	BO551605	BO551605	
C 284	14.6	81.1	545	1	AI258895	AI258895	
C 285	14.6	81.1	545	1	BI228547	BI228547	
C 286	14.6	81.1	546	1	AI549761	AI549761	
C 287	14.6	81.1	546	7	CO156796	CO156796	
C 288	14.6	81.1	551	4	AM943954	AM943954	
C 289	14.6	81.1	551	4	BI640684	BI640684	
C 290	14.6	81.1	552	7	CF246273	CF246273	
C 291	14.6	81.1	553	4	BI168680	BI168680	
C 292	14.6	81.1	558	7	CF945234	CF945234	
C 293	14.6	81.1	560	7	CF737916	CF737916	
C 294	14.6	81.1	562	2	BE144120	BE144120	
C 295	14.6	81.1	562	7	CF247768	CF247768	
C 296	14.6	81.1	563	8	BI2394071	BI2394071	
C 297	14.6	81.1	568	4	BI237871	BI237871	
C 298	14.6	81.1	569	4	BG974491	BG974491	
C 299	14.6	81.1	572	4	BG578433	BG578433	
C 300	14.6	81.1	573	7	CF863689	CF863689	

## ALIGNMENTS

RESULT 1  
CR810834/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CR810834  
GR0AA34BH01FM1 INRA BAC Bos taurus genomic clone INRA\_586B08, DNA  
sequence, genomic survey sequence.  
CR810834  
CR810834.1 GI:52692005  
GSS.  
Bos taurus (cow)  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
location/Qualifiers  
1. 742  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/strain="Breed: Holstein"  
/db\_xref="taxon:9913"  
/clone="INRA\_586B08"  
/sex="Male"  
/cell\_type="fibroblast"  
/clone\_lib="INRA bovine BAC"  
/note="Vector: pBeloBAC11; Site 1: HindIII; Holstein bull;  
INRA Bovine BAC library (Male) produced by Andre  
Eggen-Genoscope sequence ID : GR0AA34BH01FM1"

Query Match 95.6%; Score 17.2; DB 9; Length 742;  
Best local Similarity 88.9%; Pred. No. 6.5e+02;  
Matches 16; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCACGTSCCGGCATG 18  
Db 287 GACCACGTSCCGGCATG 270  
|||||:|||||  
|||:|||||

RESULT 2  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

BI634046  
376 bp mRNA linear EST 10-SEP-2001  
SD28283.5, spripte SD Drosophila melanogaster Schneider L2 cell culture  
PORT2 Drosophila melanogaster cDNA clone SD28283 5 similar to  
Pro26: PBan0004097 GO: [208 core proteasome (GO:0005839);  
multicatalytic endopeptidase (GO:0004299); 208 core proteasome  
(GO:0005839); multicatalytic endopeptidase (GO:0004299)] located  
on: 3L 73A10-73A10; 05/24/2001, mRNA sequence.

BI634046  
BI634046.1 GI:15536256  
EST.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 376)  
Harvey, D., Broksstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
Lewis, S., and Rubin, G. M.  
BDGP/HMMI Drosophila EST Project  
Unpublished (2001)

COMMENT Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST/estfruitfly.berkeley.edu>  
hit genomic AB003526: arm:3L [16482650,16763536]  
estimated-cyto:73A8-73D1: 05/24/2001  
Plate: SD,282 row: G column: 11  
High quality sequence stop: 276.  
Location/Qualifiers

FEATURES  
source  
1..376  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:7227"  
/clone="SD28283"  
/lab\_host="DHS-alpha"  
/clone\_lib="SD Drosophila melanogaster Schneider L2 cell culture POT2"  
/note="Vector: POT2; Site\_1: EcoRI; Site\_2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."

ORIGIN

Query Match 90.0%; Score 16.2; DB 4; Length 376;  
Best Local Similarity 88.2%; Pred. No. 2e+03; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACCAGTSCCGGCATG 18  
|||||:|||||  
Db 101 ACCAGTCCCGGCATG 117

RESULT 3  
CB016461  
LOCUS 703 bp mRNA linear EST 10-JAN-2003  
DEFINITION pgnic.pK012.h12 Chicken lymphoid cDNA library (pgnic) Gallus gallus  
CDNA clone pgnic.pK012.h12 5' similar to gb|AAH27139.1 similar to  
prion protein interacting protein [Mus musculus], mRNA sequence.  
CB016461  
ACCESSION CB016461.1 GI:27591197  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 703)  
Morgan, R.W. and Burnside, J.  
Chicken ESTs from lymphoid tissue  
Unpublished (2002)  
Contact: Robin W. Morgan  
University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1341  
Fax: 302-831-2822  
Email: [morganudel.edu](mailto:morganudel.edu), [www.chickest.udel.edu](http://www.chickest.udel.edu).

FEATURES  
source  
1..703  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone="pgnic.pK012.h12"  
/sex="Male and Female"  
/tissue\_type="thymus, bursa, spleen, PBL, bone marrow"  
/lab\_host="E.coli EMD108"  
/clone\_lib="Chicken lymphoid cDNA library (pgnic)"  
/note="Vector: pCMVSPORT 6"

ORIGIN

Query Match 90.0%; Score 16.2; DB 6; Length 703;  
Best Local Similarity 88.2%; Pred. No. 2e+03; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACCAGTSCCGGCATG 18  
|||||:|||||  
Db 496 ACCAGTCCCGGCATG 512

RESULT 4  
CN225729  
LOCUS 738 bp mRNA linear EST 09-APR-2004  
DEFINITION WIA077C07.abi W1brain Gallus gallus cDNA 5', mRNA sequence.  
CN225729  
ACCESSION CN225729  
VERSION CN225729.1 GI:46329220  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 738)  
Savolainen, P., Fitzsimmons, C.J., Arvestad, L., Andersson, L. and  
Lundeberg, J.  
EST analysis of brain and testis cDNA libraries from White Leghorn  
and Red Jungle Fowl  
Unpublished (2004)  
Contact: Peter Savolainen  
Department of Biotechnology  
Royal Institute of Technology, KTH  
SE-106 91 Stockholm, SWEDEN  
Tel: +46 (0)8 5537 8481  
Fax: +46 (0)8 5537 8335  
Email: [Peter.Savolainen@biotech.kth.se](mailto:Peter.Savolainen@biotech.kth.se)  
Seq primer: M13 reverse primer.  
Location/Qualifiers

FEATURES  
source  
1..738  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn"  
/db\_xref="taxon:9031"  
/sex="female"  
/lab\_host="E1ectromax DH10B (invitrogen)"  
/clone\_lib="W1brain"  
/note="Organ: brain; Vector: pSPORT-1; Site\_1: Hind III;  
Site\_2: EcoRI; The cDNA libraries were created with the  
Superscript Plasmid System (Invitrogen)."

ORIGIN

Query Match 90.0%; Score 16.2; DB 7; Length 738;  
Best Local Similarity 86.2%; Pred. No. 2e+03; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACCAGTSCCGGCATG 18  
|||||:|||||  
Db 418 ACCAGTCCCGGCATG 434

RESULT 5  
BF813258  
LOCUS 223 bp mRNA linear EST 12-JAN-2001  
DEFINITION PMA-C10113-251100-005-d06 C10113 Homo sapiens cDNA, mRNA sequence.  
BF813258  
ACCESSION BF813258.1 GI:12144019  
VERSION BF813258.1 GI:12144019  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 223)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Negai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags					
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)					
MEDLINE	20202863					
PUBMED	10737800					
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente,109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the PAPSP/RICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l=Ct10113-251100-005-d06&ct=2000-11-25&cl=1) Seq primer: puc 18 forward High quality sequence start: 16 Location/Qualifiers					
FEATURES	1..223					
Source	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_idb="Ct10113" /note="Organ: colon,ins; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."					
ORIGIN						
Query Match	86.7%	Score 15.6;	DB 2;	Length 223;		
Best Local Similarity	83.3%	Prod. NO. 4e+03;				
Matches 15;	Conservative 3;	Mismatches 1;	Indels 0;	Gaps 0;		
Oy	1 GACCACGTCGCCGGCATG 18      :::  :     145 GACCAAGTCCCGGCATG 128					
Dn						
RESULT 6	BB397065 293 bp mRNA linear EST 15-JUN-2000					
LOCUS	BB397065 RIKEN full-length enriched, ES cells Mus musculus CDNA					
DEFINITION	Clone C330002B21.3', mRNA sequence.					
ACCESSION	BB397065					
VERSION	BB397065.1 GI:9216461					
KEYWORDS	EST.					
SOURCE	Mus musculus (house mouse)					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 293)					
AUTHORS	Kono, H., Atzawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carinci, P., Endo, T., Fukuda, S., Fukuishi, Y., Hara, A., Hayatsu, N., Hitozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Okazaki, M., Matsuyama, T., Miura, R., Mizuno, Y., Nakamura, M., Oda, H., Otsuki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigenoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamuro, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiaki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.					
TITLE	RIKEN Mouse ESTs (Kono, H., et al.)					
JOURNAL	Unpublished (2000)					

**COMMENT**

Contact: Yoshinide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-resgsc.riken.jp, URL: http://genome-gsc.riken.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasakura, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotranscription and thermoadaptation of thermophilic enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Ito, M., Katsunai, T., Akizawa, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shiba, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

**FEATURES**

**SOURCE**

1. 293

location/Qualifiers

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="CG30002B21"

/cell\_type="ES cells"

/lab\_host="SOLR"

/note="Site 1: XhoI; Site 2: BamHI; CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5].  
GAGGAGAGAGATCTCGATTAAATTAATCATCCCCCCCCCCC 3'." GAGGAGAGAGATCTCGATTAAATTAATCATCCCCCCCCCCC 3'."

**ORIGIN**

Query Match      86.7%; Score 15.6; DB 2; Length 293;  
Best Local Similarity    83.3%; Pred. No. 4e+03;  
Matches    15; Conservative    2; Mismatches    1; Indels    0; Gaps    0;

```

Ox      1  GACCACTGCCGGCATG  18
          |||||::|||:||||
Db      44 GACCACGTCCCGGCATG  61

```

**RESULT 7**

CB995647                  295 bp    mRNA    linear    EST 01-MAY-2003

LOCUS                    CB995647

DEFINITION             AGNCOCURT 13617193 NIH\_MGC\_148 Homo sapiens cDNA clone IMAGE:30339972 5', mRNA sequence.

ACCESSION               CB995647

VERSION                  CB995647.1 GI:30290167

KEYWORDS                EST.

SOURCE                  Homo sapiens (human)

ORGANISM                Homo sapiens

Eukaryote; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE               1 (bases 1 to 295)

TITLE                    NIH-MGC <http://mgc.ncl.nih.gov/>. National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-rmail.nih.gov  
Tissue Procurement: Dr. Stefan Hansson  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help and advice from Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at:  
http://image.llnl.gov  
Plate: NDAM369 row: b column: 13  
High quality sequence stop: 295.  
Location/Qualifiers  
1..295  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30339972"  
/tissue\_type="pre-ecclampic placenta"  
/lab\_host="DH10B TONa"  
/clone\_lib="NIH MGC 148"  
/note="Organ: Placenta; Vector: pBluescriptR; Site 1: all-xhoI; Site 2: BamH; Library is oligo-dT primed and directionally cloned using primer 5'-TTTTTCTTTTCTTCTT-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH MGC library."

ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 295;  
Best Local Similarity 83.3%; Pred. No. 4e+03; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACGTCSCGCGCATG 18  
|||||:|:|:|:|:|:|  
171 GACCACGCGCCGCGCATG 188

Db

RESULT 8  
BF704768/c 433 bp mRNA linear EST 29-DEC-2000  
LOCUS F01.51.C08.b1.A003 Floral-Induced Meristem 1 (F01) Sorghum  
DEFINITION F01.51.C08.b1.A003 Floral-Induced Meristem 1 (F01) Sorghum  
proingnum cDNA, mRNA sequence.  
BF704768  
ACCESSION BF704768.1 GI:11996429  
VERSION  
KEYWORDS EST.  
SOURCE Sorghum propinquum  
ORGANISM Sorghum propinquum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 433)  
Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt, L.H.  
An EST database from Sorghum: floral-induced meristems  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
Seq primer: JEN REV  
High quality sequence stop: 429

TITLE  
JOURNAL  
COMMENT

FEATURES  
source

POLYA=No.  
Location/Qualifiers  
1..433  
/organism="Sorghum propinquum"  
/mol\_type="mRNA"  
/db\_xref="taxon:132711"  
/clone\_lib="Floral-Induced Meristem 1 (F01)"  
/note="Organ: Floral-induced meristems; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

ORIGIN  
Query Match 86.7%; Score 15.6; DB 2; Length 433;  
Best Local Similarity 83.3%; Pred. No. 4e+03; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACGTCSCGCGCATG 18  
|||||:|:|:|:|:|:|  
334 GCCACGTCGCCGCGCATG 317

Db

RESULT 9  
BF656537/c 435 bp mRNA linear EST 20-DEC-2000  
LOCUS F01.51.C08.g1.A003 Floral-Induced Meristem 1 (F01) Sorghum  
DEFINITION F01.51.C08.g1.A003 Floral-Induced Meristem 1 (F01) Sorghum  
proingnum cDNA, mRNA sequence.  
BF656537  
ACCESSION BF656537.1 GI:11921671  
VERSION  
KEYWORDS EST.  
SOURCE Sorghum propinquum  
ORGANISM Sorghum propinquum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 435)  
Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt, L.H.  
An EST database from Sorghum: floral-induced meristems  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
Seq primer: T7  
High quality sequence start: 22  
High quality sequence stop: 435  
POLYA=Yes.  
Location/Qualifiers  
1..435  
/organism="Sorghum propinquum"  
/mol\_type="mRNA"  
/db\_xref="taxon:132711"  
/clone\_lib="Floral-Induced Meristem 1 (F01)"  
/note="Organ: Floral-induced meristems; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The

TITLE  
JOURNAL  
COMMENT

FEATURES  
source

## ORIGIN

library was made from poly-A RNA in the cloning vector  
lambda ZAP II. Clones to be sequenced were prepared by  
mass excision."

Query Match 86.7%; Score 15.6; DB 2; Length 435;  
Best Local Similarity 83.3%; Pred. No. 4e+03; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 GACCACGTSCCGGCATG 18  
|||||:|||||  
204 GCCCAGTCCCGGCATG 187

RESULT 10  
BB729006/c  
LOCUS BB729006 451 bp mRNA linear EST 12-OCT-2001  
DEFINITION BB729006 RIKEN full-length enriched, 8 cells embryo Mus musculus  
CDS clone BB60104115 3', mRNA sequence.  
BB729006  
BB729006.1 GI:16112281  
EST.  
Mus musculus (house mouse)  
SOURCE  
ORGANISM  
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 451)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hangaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K.,  
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Komoto,H., Kouda,M.,  
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,  
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaki-Akahira,S.,  
Tanaka,T., Tomaru,A., Toya,T., Watanabe,A., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

TITLE Unpublished (2001)  
JOURNAL Contact: Yoshihide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Komoto,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,  
Matsuyama,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.  
and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Komoto,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,  
Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.

## FEATURES

Location/Qualifiers  
1..451  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"

## ORIGIN

/clone="BB60104115"  
/cell\_type="8 cells embryo"  
/dev\_stage="8 cells embryo"  
/clone\_lib="RIKEN full-length enriched, 8 cells embryo"

Query Match 86.7%; Score 15.6; DB 2; Length 451;  
Best Local Similarity 83.3%; Pred. No. 4e+03; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 GACCACGTSCCGGCATG 18  
|||||:|||||  
28 GCCCAGTCCCGGCATG 11

RESULT 11  
CL797916  
LOCUS CL797916 510 bp DNA linear GSS 06-AUG-2004  
DEFINITION OR\_CBA0008006.r OR\_CBA Oryza rufipogon genomic clone OR\_CBA0008006  
3', genomic survey sequence.  
CL797916  
CL797916.1 GI:51020275  
GSS.  
SOURCE  
ORGANISM  
Oryza rufipogon  
Oryza rufipogon  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 510)  
Kim,H., Yu,L., Wisotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,  
Jettly,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and  
Wing,R.  
OMAP project  
Unpublished (2004)  
Contact: Rod A. Wing  
Arizona Genomics Institute  
University of Arizona  
Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 1259  
Email: http://genome.arizona.edu

REFERENCE PCR Primers  
FORWARD: TAA TAC GAC TCA CTA TAG GG  
BACKWARD: CAC TCA TTA GGC ACC CCA  
Plate: 0008 row: J column: 06  
Seq primer: CAC TCA TTA GGC ACC CCA  
Class: BAC ends.

## FEATURES

Location/Qualifiers  
1..510  
/organism="Oryza rufipogon"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4529"  
/clone="OR\_CBA0008006"  
/tissue\_type="young leaves"  
/dev\_stage="2 week old seedlings"  
/lab\_host="DH10B T1 phage resistant"  
/clone\_lib="OR\_CBA"  
/note="Vector: pAG1BAC1; Site 1: HindIII; Site 2: HindIII;  
dir treated 36 hrs before harvest"

## ORIGIN

Query Match 86.7%; Score 15.6; DB 9; Length 510;  
Best Local Similarity 83.3%; Pred. No. 4e+03; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 GACCACGTSCCGGCATG 18  
|||||:|||||  
335 GACCACGTGCGCGGCATG 352

RESULT 12  
AQ397255/c  
LOCUS AQ397255 514 bp DNA linear GSS 06-MAR-1999

COMMENT	JOURNAL TITLE Xenelaxis FlyTag EST Project CK02 Library Unpublished (2004) Contact: Stapleton, M. BDGP Lawrence Berkeley National Lab One Cyclotron Rd, Berkeley, CA 94720, USA Fax: 510 486 6798 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu Plate: EP.185 row: A column: 12 High quality sequence stop: 401.
FEATURES	source 1..523 /organism="Drosophila melanogaster" /mol_type="mRNA" /db_xref="taxon:7227" /clone_id="EP18512" /clone_1id="Xenelaxis FlyTag CK02 pCDNA-SK+ " /note="Organ: mixed stage embryos, imaginal disks, and adult heads; Vector: pCDNA-SK+; Site 1: NotI; Site 2: XhoI; Random primed, normalized library from mixed stage embryos, imaginal disks, and adult heads. Subset of Xenelaxis FlyTag CK01 clones sequenced from 3' end"
ORIGIN	Query Match 86.7%; Score 15.6; DB 7; Length 523; Best Local Similarity 83.3%; Pred. No. 4e+03; Indels 0; Gaps 0; Matches 15; Conservative 2; Mismatches 1;
QY	1 GACCAAGTSCCGGCGATG 18      :::      255 GTCCACGTGCCGGCATG 238
Dn	
RESULT 14	AL903319 539 bp mRNA linear EST 02-JUN-2004
LOCUS	AL903319
ACCSSION	AL903319 PUB-Z1+Z2 Danio rerio cDNA clone 001-B06-1, mRNA sequence.
VERSION	AL903319.1 GI:23165878
KEYWORDS	EST.
SOURCE	Danio rerio (zebrafish)
ORGANISM	Danio rerio Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE	1 (bases 1 to 539) Lo,J., Lee,S., Xu,M., Liu,F., Ruan,H., Eun,A., He,Y., Ma,W., Wang,W., Wen,Z. and Peng,J. 15000 unique zebrafish EST clusters and their future use in microarray for profiling gene expression patterns during embryogenesis
AUTHORS	Genome Res. 13 (3), 455-466 (2003)
TITLE	Contact: Peng J Lab of Functional Genomics Institute of Molecular and Cell Biology 30 Medical Drive, Singapore, 117609, Singapore Email: pengji@mcb.a-star.edu.sg Clone requests: info@openbiosystems.com Open Biosystems, 6705 Odysseey Drive, Huntsville, AL 35806.
JOURNAL MEDLINE	location/Qualifiers
PUBMED COMMENT	1..539 /organism="Danio rerio" /mol_type="mRNA" /strain="local wildtype" /db_xref="taxon:7955" /clone="001-B06-1" /tissue type="whole embryo or fish" /dev_stage="mixed stages" /clone_1id="PUB-Z1+Z2"
FEATURES	
source	
ORIGIN	



Query Match 86.7%; Score 15.6; DB 1; Length 539;  
Best Local Similarity 83.3%; Pred. No. 4e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
Db 260 GACCACGTCCGGGCTTG 277

RESULT 15  
LBAFB3C09/C 564 bp DNA linear GSS 20-JUN-2003  
Leishmania braziliensis GSS, clone LBAFB3C09, genomic survey  
Sequence.  
BX544274  
BX544274.1 GI:32139513  
GSS, genomic survey sequence.  
Leishmania braziliensis  
Leishmania braziliensis  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Leishmania; Leishmania braziliensis species complex.

REFERENCE 1  
AUTHORS Laurentino, E.C., Ruiz, J.C. and Cruz, A.K.  
TITLE GSS analysis of the Leishmania braziliensis genome  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 564)  
AUTHORS Cruz, A.K.  
TITLE Direct Submission  
SUBMITTED (17-JUN-2003) Cruz, A.K., University of Sao Paulo,  
Department of Molecular and Cell Biology, FMRP, Avenida  
Bandeiraantes 3900, Ribeirao Preto, SP, 14049-900, BRAZIL  
Clone requests: akcruz@fmrp.usp.br.

COMMENT  
FEATURES  
SOURCE location/Qualifiers  
1. 564  
/organism="Leishmania braziliensis"  
/mol\_type="genomic DNA"  
/strain="MHOM/BR/75/M2904"  
/db\_xref="taxon:5660"  
/clone="LBAFB3C09"

ORIGIN  
Query Match 86.7%; Score 15.6; DB 9; Length 564;  
Best Local Similarity 83.3%; Pred. No. 4e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
Db 355 GACCACGTCCGGGCATG 338

RESULT 16  
BP221909 581 bp mRNA linear EST 15-SEP-2004  
BP221909 Sugano cDNA library, COL0679 Homo sapiens cDNA clone  
COR00491, mRNA sequence.  
BP221909  
BP221909.1 GI:52094814  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 581)  
Suzuki, Y., Yamashita, R., Shiota, M., Sakakibara, Y., Chida, J.,  
Mizushima-Sugano, J., Nakai, K. and Sugano, S.  
Sequence comparison of human and mouse genes reveals a homologous  
block structure in the promoter regions  
Genome Res. 14 (9), 1711-1718 (2004)  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

JOURNAL  
TITLES  
AUTHORS  
CONTACT  
KEYWORDS

Email: yenzuki@ims.u-tokyo.ac.jp.

FEATURES  
SOURCE location/Qualifiers  
1. 581  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="COR00491"  
/cell\_line="COL0679"  
/clone\_lib="Sugano cDNA library, COL0679"  
/note="melanoma"

ORIGIN  
Query Match 86.7%; Score 15.6; DB 5; Length 581;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
Db 117 GACCACGGCCCGGCATG 134

RESULT 17  
BP356698 582 bp mRNA linear EST 17-SEP-2004  
BP356698 Sugano cDNA library, mammary gland T47D Homo sapiens cDNA  
clone TDR01021, mRNA sequence.  
BP356698  
BP356698.1 GI:52286705  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 582)  
Suzuki, Y., Yamashita, R., Shiota, M., Sakakibara, Y., Chida, J.,  
Mizushima-Sugano, J., Nakai, K. and Sugano, S.  
Sequence comparison of human and mouse genes reveals a homologous  
block structure in the promoter regions  
Genome Res. 14 (9), 1711-1718 (2004)  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yenzuki@ims.u-tokyo.ac.jp.

COMMENT  
FEATURES  
SOURCE location/Qualifiers  
1. 582  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TDR01021"  
/issue\_type="mammary gland"  
/cell\_line="T47D"  
/clone\_lib="Sugano cDNA library, mammary gland T47D"  
/note="mammary gland tumor"

ORIGIN  
Query Match 86.7%; Score 15.6; DB 5; Length 582;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
Db 163 GACCACGGCCCGGCATG 180

RESULT 18  
BP364560 582 bp mRNA linear EST 17-SEP-2004  
BP364560 Sugano cDNA library, fetal lung fibroblast TIG Homo  
sapiens cDNA clone TIR04501, mRNA sequence.  
BP364560  
BP364560.1 GI:52294802  
EST.

JOURNAL  
TITLES  
AUTHORS  
CONTACT  
KEYWORDS

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 582)  
Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J., Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
TITLE Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions  
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp.  
Location/Qualifiers

FEATURES  
source  
1..582  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TIR04501"  
/tissue\_type="lung"  
/cell\_type="fibroblast"  
/cell\_line="TIG"  
/dev\_stage="fetal"  
/clone\_lib="Sugano cDNA library, fetal lung fibroblast TIG"

ORIGIN  
Query Match 86.7%; Score 15.6; DB 5; Length 582;  
Best Local Similarity 83.3%; Pred.No.4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTCSCGGCATG 18  
||||| :||:|||||  
144 GACCACGGCCCGGCATG 161

DB  
BP280394 583 bp mRNA linear EST 16-SEP-2004  
BP280394 Sugano cDNA library, KG-1-C Homo sapiens cDNA clone  
LOCUS BP280394  
DEFINITION BP280394, mRNA sequence.  
ACCESSION BP280394  
VERSION BP280394.1 GI:52194126  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 583)  
Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J., Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
TITLE Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions  
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp.  
Location/Qualifiers

FEATURES  
source  
1..583  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="KGR07789"  
/cell\_line="KG-1-C"  
/clone\_lib="Sugano cDNA library, KG-1-C"  
/note="glioma"

ORIGIN  
Query Match 86.7%; Score 15.6; DB 5; Length 583;  
Best Local Similarity 83.3%; Pred.No.4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Best Local Similarity 83.3%; Pred.No.4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTCSCGGCATG 18  
||||| :||:|||||  
136 GACCACGGCCCGGCATG 153

DB  
BP359091 583 bp mRNA linear EST 17-SEP-2004  
BP359091 Sugano cDNA library, mammary gland T47D Homo sapiens cDNA  
LOCUS BP359091  
DEFINITION BP359091, mRNA sequence.  
ACCESSION BP359091  
VERSION BP359091.1 GI:52289104  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 583)  
Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J., Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
TITLE Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions  
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp.  
Location/Qualifiers

FEATURES  
source  
1..583  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TDR07864"  
/tissue\_type="mammary gland"  
/cell\_line="T47D"  
/clone\_lib="Sugano cDNA library, mammary gland T47D"  
/note="mammary gland tumor"

ORIGIN  
Query Match 86.7%; Score 15.6; DB 5; Length 583;  
Best Local Similarity 83.3%; Pred.No.4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTCSCGGCATG 18  
||||| :||:|||||  
136 GACCACGGCCCGGCATG 153

DB  
BP364824 587 bp mRNA linear EST 17-SEP-2004  
BP364824 Sugano cDNA library, fetal lung fibroblast TIG Homo sapiens cDNA clone TIR05159, mRNA sequence.  
LOCUS BP364824  
DEFINITION BP364824, mRNA sequence.  
ACCESSION BP364824  
VERSION BP364824.1 GI:52295066  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 587)  
Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J., Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
TITLE Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions  
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo

4-6-1, Shiokaneda, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp.

## FEATURES

SOURCE

1. 587  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TIR05159"  
/tissue\_type="lung"  
/cell\_type="fibroblast"  
/cell\_line="RIG"  
/dev\_stage="fetal"  
/clone\_lib="Sugano CDNA library, fetal lung fibroblast  
TIG"

## ORIGIN

Query Match 86.7%; Score 15.6; DB 5; Length 587;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAGTSCCGGCGCATG 18  
|||||:|||||  
26 GACCAGCGCCCGGCGCATG 43

RESULT 22  
BP310516 588 bp mRNA linear EST 17-SEP-2004  
LOCUS BP310516  
DEFINITION BP310516 Sugano CDNA library, brain Homo sapiens CDNA clone  
ACCESSION NR02745, mRNA sequence.  
VERSION BP310516  
KEYWORDS BP310516.1 GI:52239491  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 588)  
Suzuki, Y., Yamashita, R., Shiota, M., Sakakibara, Y., Chida, J.,  
Mizushima-Sugano, J., Nakai, K. and Sugano, S.  
Sequence comparison of human and mouse genes reveals a homologous  
block structure in the promoter regions  
Genome Res. 14 (9), 1711-1718 (2004)  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shiokaneda, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp.

TITLE  
JOURNAL  
COMMENT

## FEATURES

SOURCE

1. 588  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NR02745"  
/tissue\_type="brain"  
/clone\_lib="Sugano CDNA library, brain"

## ORIGIN

Query Match 86.7%; Score 15.6; DB 5; Length 588;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAGTSCCGGCGCATG 18  
|||||:|||||  
136 GACCAGCGCCCGGCGCATG 153

RESULT 23  
AZ411523 588 bp DNA linear GSS 03-OCT-2000  
LOCUS AZ411523  
DEFINITION IM0184C11R Mouse 10kb plasmid U06C1M library Mus musculus genomic  
clone U06C1M0184C11 R, genomic survey sequence.  
ACCESSION AZ411523

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AZ411523.1 GI:10535536  
GSS.  
Mus musculus (house mouse)  
Mus musculus

## REFERENCE

1 (bases 1 to 588)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,  
Rielly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von  
Niederhausern, A. and Wright, D., Weis, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)

## TITLE

plasmid inserts

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Robert B. Weis  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0184 row: C column: 11  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 588.  
Location/Qualifiers

## FEATURES

SOURCE

1. 588  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="U06C1M0184C11"  
/sex="Male"  
/lib\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid U06C1M library"  
/note="Vector: pMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD42 (g14732114[gb|AF129072.1]), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptor complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

## ORIGIN

Query Match 86.7%; Score 15.6; DB 8; Length 588;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAGTSCCGGCGCATG 18  
|||||:|||||  
575 GACCAGGTGCGGCGCATG 558

RESULT 24  
A1389668 602 bp mRNA linear EST 02-DEC-2003  
LOCUS A1389668  
DEFINITION GH20904.5prtime GH Drosophila melanogaster head P072 Drosophila  
melanogaster cDNA clone GH20904 5 similar to CG6305; Fpan0006305  
GO: [structural protein (GO:0005198)] located on: 2R 50C9-50C11;

08/13/2002, mRNA sequence.  
 AI389668  
 VERSION AI389668.1 GI:4203679  
 EST.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster (fruit fly)  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 602)  
 REFERENCE  
 AUTHORS Harvey D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S., and Rubin, G.M.  
 TITLE BDGP/HMT Drosophila EST Project  
 JOURNAL Unpublished (2001)  
 COMMENT Other ESTs: GH20904.3prime  
 Contact: Stapleton, M.  
 BDBG  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: <http://www.fruitfly.org/EST/est@fruitfly.berkeley.edu>  
 hit genomic AB003817: arm:2R [8578716,8840972]  
 estimated-cyto:50B9-50C14: 04/10/2001  
 plate: GH-209 row: A column: 4  
 High quality sequence stop: 509  
 POLYA=No.  
 FEATURES  
 source Location/Qualifiers  
 1..602  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7227"  
 /clone="GH20904"  
 /sex="male and female"  
 /dev\_stage="adult"  
 /lab\_host="DH5 - alpha"  
 /clone\_lib="GH Drosophila melanogaster head POT2"  
 /note="Organ: head; Vector: POT2; Site: 1: EcoRI; Site: 2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."  
 ORIGIN  
 Query Match 86.7%; Score 15.6; DB 1; Length 602;  
 Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GACCACGTCGCGCATG 18  
 Db 454 GTCACGTCGCCGCGCATG 471  
 RESULT 25 612 bp mRNA linear EST 16-SEP-2003  
 CD311022/c Strp691.004386 Sea urchin larva cDNA library MPMG691  
 LOCUS Strongylocentrotus purpuratus cDNA clone  
 MPMG691H0297;MPI\_SURUDI\_97H2 5', mRNA sequence.  
 DEFINITION CD311022  
 VERSION CD311022  
 KEYWORDS EST.  
 SOURCE Strongylocentrotus purpuratus  
 ORGANISM Strongylocentrotus purpuratus  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinidea; Echinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.  
 1 (bases 1 to 612)  
 REFERENCE  
 AUTHORS Pouletka, A.J., Groth, D., Hennig, S., Thamm, S., Cameron, A., Beck, A., Reinhardt, R., Herwig, R., Panopoulou, G., and Lehrach, H.  
 TITLE Generation, annotation, evolutionary analysis, and database integration of 20,000 unique sea urchin EST clusters  
 JOURNAL Genome Res. 13 (12), 2736-2746 (2003)  
 COMMENT Contact: Pouletka AJ  
 Laboratory 145, dept. Lehrach  
 Max-Planck-Institut fuer Molekulare Genetik

Imhestr.63-73, D-14195 Berlin, Germany  
 Tel.: +49 30 8413 1235  
 Fax: +49 30 8413 1128  
 Email: [pouletka@molgen.mpg.de](mailto:pouletka@molgen.mpg.de)  
 The library was characterised by oligonucleotide fingerprinting (ONF) to reduce sequencing redundancy. According to the ONF procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONF cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the other clones assigned to the same ONF cluster as the clone from which the above EST is generated is available at the sea urchin project web site at: [http://www.molgen.mpg.de/ag\\_searchin/](http://www.molgen.mpg.de/ag_searchin/). cDNA clones and fillers are distributed via the Resource Center/Primary Database of the German Human Genome Project (<http://www.rzpd.de>)  
 PCR Primers  
 FORWARD: 5' CCCAGGCTTACCTTATGCTCCGGCTCG 3' (M13RSP) 5'-seq  
 BACKWARD: 5' GCTATTACGCACTGCGCAAGGGGATGTC 3' (M13FSP) 3'-seq  
 Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSPORT3/86  
 High quality sequence stop: 612.  
 FEATURES  
 source Location/Qualifiers  
 1..612  
 /organism="Strongylocentrotus purpuratus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7668"  
 /clone="MPMG691H0297;MPI\_SURUDI\_97H2"  
 /tissue\_type="whole larva"  
 /dev\_stage="larva 2-3 weeks"  
 /lab\_host="E. coli, XLI blue"  
 /clone\_lib="Sea urchin larva cDNA library MPMG691"  
 /note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; Random primed and directionally cloned in pSPORT1 vector using a NotI (5'-TGACTAGTCTGATCGGATCGGCGCGCC (T)15-3' and a SalI 5'-TCGACCAACGCTCCG-3' adapters (Gibco BRL)."  
 ORIGIN  
 Query Match 86.7%; Score 15.6; DB 6; Length 612;  
 Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GACCACGTCGCGCATG 18  
 Db 321 GACCACGTCGCGCATG 304  
 RESULT 26 624 bp mRNA linear EST 18-MAY-2004  
 CN720685 E0808A04-5 NIA Mouse four-cell-embryo cDNA library (long) Mus  
 LOCUS musculus cDNA clone NIA:E0808A04 IMAGE:30906339 5', mRNA sequence.  
 DEFINITION CN720685  
 VERSION CN720685  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 624)  
 REFERENCE  
 AUTHORS Sharov, A.A., Piao, Y., Matoba, R., Dudekula, D.B., Qian, Y., Vanburen, V., Falco, G., Martin, P.R., Stagg, C.A., Basse, V.C., Wang, Y., Carter, M.G., Hamatani, T., Alba, K., Akutsu, H., Sharova, L., Tanaka, T.S., Kimber, W.L., Yoshikawa, T., Jaradat, S.A., Pantano, S., Nagaraja, R., Boheler, K.R., Taub, D., Hodas, R.J., Longo, D.L., Schlesinger, D., Keller, J., Klotz, E., Kelsae, G., Umezawa, A., Vesecov, A.L., Rossant, J., Kumachi, T., Hogan, B.L., Curci, A., D'Urso, M., Kelsae, J., Hide, W., and Ko, M.S.  
 TITLE Transcriptional analysis of mouse stem cells and early embryos  
 JOURNAL PLOS Biol. 1 (3), 410-419 (2003)  
 COMMENT Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov  
 Plate: E0808 Igsun: A Column: 04  
 Seq primer: M13 Reverse  
 High quality sequence stop: 624  
 POLYA=NO.

# FEATURES

source

1. 624  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:E0808A04-5"  
 /db\_xref="taxon:10090"  
 /clone="NIA:E0808A04 IMAGE:3096339"  
 /issue\_type="4-cell stage embryo"  
 /dev\_stage="4-cell"  
 /lab\_host="DH10B"  
 /clone\_1b="NIA Mouse four-cell-Embryo cDNA library (long)"  
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site\_1: SalI; Site\_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://igsun.grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). The mRNAs were extracted from a pool of 360 embryos at 4-cell stage. Double-stranded cDNAs were synthesized with an Oligo(dt) primer (Invitrogen): 5'-TGACTAGTTCATGATCGGAGCGCCGCTTTTCTTTT-3' from 10-8ng of mRNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2Kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 86.7%; Score 15.6; DB 7; Length 624;  
 Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACTGSCCGGCATG 18  
 |||||:|||||  
 Db 100 GACCACGTGCGGCGCATG 117

## RESULT 27

CA100015 631 bp mRNA linear EST 23-SEP-2003  
 LOCUS SCVPC16044C07.G Cl6 Saccharum officinarum cDNA clone SCVPC16044C07  
 DEFINITION 5', mRNA sequence.  
 ACCESSION CA100015  
 VERSION CA100015.1 GI:34953322  
 KEYWORDS EST.  
 SOURCE Saccharum officinarum  
 ORGANISM Saccharum officinarum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum  
 complex.

REFERENCE 1 (bases 1 to 631)  
 Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.  
 The libraries that made SUCEST  
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)  
 CONTACT: Arruda P  
 Centro de Biologia Molecular e Engenharia Genetica  
 Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
 Tel: 55 19 3788 1137  
 Fax: 55 19 3788 1089  
 Email: parreda@unicamp.br  
 Clone distribution information can be found  
 through the Brazilian Clone Collection Center (BCCC) at  
<http://www.bcccenter.fcav.unesp.br>  
 Plate: 044 row: C column: 07  
 Seq primer: T7 Promoter Primer.

# FEATURES

source

1. 631  
 Location/Qualifiers  
 /organism="Saccharum officinarum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4547"  
 /clone="SCVPC16044C07"  
 /lab\_host="DH10B"  
 /clone\_1b="Cl6"  
 /note="Organ: Pool of sugarcane calli submitted to low (40C) and high (37 C) temperature stress; Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; An unidirectional cDNA library generated from (pool of sugarcane calli submitted to low (40C) and high (37 C) temperature stress). cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-column and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://sucest.lad.ic.unicamp.br/public>"

## ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 631;  
 Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACTGSCCGGCATG 18  
 |||||:|||||  
 Db 300 GACCACGTGCGGCGCATG 317

## RESULT 28

CA071958 634 bp mRNA linear EST 23-SEP-2003  
 LOCUS SCBGAM1092E04.G AM1 Saccharum officinarum cDNA clone SCBGAM1092E04  
 DEFINITION 5', mRNA sequence.  
 ACCESSION CA071958  
 VERSION CA071958.1 GI:34924097  
 KEYWORDS EST.  
 SOURCE Saccharum officinarum  
 ORGANISM Saccharum officinarum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum  
 complex.

REFERENCE 1 (bases 1 to 634)  
 Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.  
 The libraries that made SUCEST  
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)  
 CONTACT: Arruda P  
 Centro de Biologia Molecular e Engenharia Genetica  
 Universidade Estadual de Campinas  
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
 Tel: 55 19 3788 1137  
 Fax: 55 19 3788 1089  
 Email: parreda@unicamp.br  
 Clone distribution information can be found  
 through the Brazilian Clone Collection Center (BCCC) at  
<http://www.bcccenter.fcav.unesp.br>  
 Plate: 092 row: E column: 04  
 Seq primer: T7 Promoter Primer.

# FEATURES

source

1. 634  
 Location/Qualifiers  
 /organism="Saccharum officinarum"

/mol\_type="mRNA"  
/db\_xref="taxon:4547"  
/clone="SCBGM1092E04"  
/lab\_host="DH10B"  
/clone\_lib="AM1"  
/note="Organ: Apical meristem and tissues surrounding of mature plants; Vector: pSport1; Site\_1: Salt; Site\_2: Not; An unidirectional cDNA library generated from [Apical meristem and tissues surrounding of mature plants]. cDNA was prepared from polyA+ mRNA using Superscript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a separose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://succeat.lad.ic.unicaamp.br/public>"

## ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 634;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAAGTSCCGGCATG 18  
|||||:|||||  
Db 168 GACCAAGCGCCGGGCATG 185

RESULT 29  
CF367459 636 bp mRNA linear EST 25-AUG-2003  
LOCUS 842208 MARC 3P1G Sus scrofa cDNA 5', mRNA sequence.  
ACCESSION CF367459  
VERSION CF367459.1 GI:34172024  
KEYWORDS EST.  
SOURCE Sus scrofa (pig)  
ORGANISM Sus scrofa (pig)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
1 (bases 1 to 636)  
Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,  
Wise,T.A., Nonneman,D.J., Wray,J.E., and Keefe,J.W.  
A second set of porcine ESTs from a pooled-tissue normalized  
library  
Unpublished (2003)  
JOURNAL  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: [smith@email.marc.usda.gov](mailto:smith@email.marc.usda.gov)  
Single pass sequencing. Bases called with phred v0.020425.c and  
trimmed with the aid of the trim\_alt option. Vector identified with  
cross match v0.990329.  
Plate: SRG8026 row: F column: 4  
Seq primer: GTATATGAGACTCATATAGCG.  
Location/Qualifiers  
1..636

FEATURES  
source  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/issue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 3P1G"  
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;  
library made with RNA pooled from multiple tissues  
including brain, liver, muscle, placenta/endometrium,  
ovary, testes, and bone marrow."

## ORIGIN

Query Match 86.7%; Score 15.6; DB 7; Length 636;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAAGTSCCGGCATG 18  
|||||:|||||  
Db 419 GACCAAGCGCCGGGCATG 402

RESULT 30  
CB096877/c 641 bp mRNA linear EST 27-JAN-2003  
LOCUS AF53-RP\_09 J17 T7 069.abl IRR1 Drought Stress Panicle Library  
DEFINITION Oryza sativa (indica cultivar-group) cDNA clone C0003305 5' similar  
to Transcription factor BTF3 (RNA polymerase B transcription factor  
3), mRNA sequence.  
CB096877  
ACCESSION CB096877.1 GI:27921065  
VERSION  
KEYWORDS EST.  
SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehretaceae; Oryzaceae; Oryza.  
1 (bases 1 to 641)  
Bennett,J., Arumugam,K., Laflitte,R., Wen,J., Rudd,S. and  
Bruskiewich,R.M.  
IRRI Drought Stress Panicle cDNA Library  
Unpublished (2002)  
JOURNAL  
COMMENT Contact: Richard Bruskiewich  
Biometrics and Bioinformatics Unit  
International Rice Research Institute  
DAPO 7777, Metro Manila, Philippines  
Tel: +63-2-845-0563  
Fax: +63-2-845-0606  
Email: [r.bruskiewich@cgiar.org](mailto:r.bruskiewich@cgiar.org)  
International Rice Information System (IRIS;  
<http://www.iris.irri.org>): D0203304 Assignment of putative function  
to the sequence by S. Rudd of the Munich Information Center for  
Protein Sequences (<http://mips.gsf.de>)  
Plate: 09 row: J column: 17.  
Location/Qualifiers  
1..641

FEATURES  
source  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="IR64"  
/db\_xref="taxon:39946"  
/clone="C0003305"  
/issue\_type="Panicles"  
/dev\_stage="Flowering"  
/clone\_lib="IRRI Drought Stress Panicle Library"  
/note="Vector: pBluescript II SK+; Water stress was  
applied by not watering for 4 consecutive days. Panicles  
were collected from control (well watered) and stressed  
plants at 2 days before heading, at heading, 50% flowering  
and 4 days after 50% flowering."

## ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 641;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAAGTSCCGGCATG 18  
|||||:|||||  
Db 499 GACCAAGTCCCGGCATG 482

Search completed: July 20, 2005, 18:15:50  
Job time : 2283.57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 13:19:40 ; Search time 296.143 Seconds  
(without alignments)  
359.811 Million cell updates/sec

Title: US-10-017-471B-1

Perfect score: 18

Sequence: 1 gaccacgtccggcagcagc 18

Scoring table: IDENTITY NUC  
Gapop 10'-0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : N\_Geneseq\_16Dec04:\*

1: geneeqn1980s:\*\n2: geneeqn1990s:\*\n3: geneeqn2000s:\*\n4: geneeqn2001as:\*\n5: geneeqn2001bs:\*\n6: geneeqn2002as:\*\n7: geneeqn2002bs:\*\n8: geneeqn2003as:\*\n9: geneeqn2003bs:\*\n10: geneeqn2003cs:\*\n11: geneeqn2003ds:\*\n12: geneeqn2004as:\*\n13: geneeqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.2	95.6	18	6	ABN86366
2	17.2	95.6	18	10	ADQ48795
3	16.2	90.0	587	13	ADQ48866
4	16	88.9	4346	6	ABN86379
5	16	88.9	4346	10	ADQ48794
6	15.6	86.7	465	3	AACT5505
7	15.6	86.7	513	2	AAAX98247
8	15.6	86.7	612	13	ADQ48795
9	15.6	86.7	744	2	ABLI10881
10	15.6	86.7	766	2	AAAX98245
11	15.6	86.7	891	11	ABDI14835
12	15.6	86.7	1188	10	ABDI14835
13	15.6	86.7	1275	11	ABDI15034
14	15.6	86.7	1408	2	AAZ32291
15	15.6	86.7	1408	2	AAZ32287
16	15.6	86.7	2000	8	ADA73004
17	15.6	86.7	2048	2	AAAX98244
18	15.6	86.7	2395	2	AAV52338
19	15.6	86.7	5042	6	ABN59696
20	15.6	86.7	5362	10	ADQ6975

21	15.6	86.7	5624	5	AAH89920	Aah89920 Human bon
22	15.6	86.7	12332	4	ABLI10880	Abli10880 Drosophila
23	15.6	86.7	110000	10	ABN56454_01	Continuation (2 of
24	15.2	84.4	5356	4	ABLI16420	Abli16420 Drosophila
25	15.2	84.4	110021	13	ABD32859	Abd32859 Mouse can
26	14.6	81.1	209	6	ABLI75000	Abli75000 Corn tass
27	14.6	81.1	520	13	ADQ52850	Adq52850 Novel can
28	14.6	81.1	521	3	AAAC56488	Aac56488 Eucalyptu
29	14.6	81.1	521	3	AAAC56107	Aac56107 Eucalyptu
30	14.6	81.1	568	9	ADA48826	Ada48826 Banana ge
31	14.6	81.1	568	11	ADJ11865	Adj11865 Banana cd
32	14.6	81.1	607	10	ADP33617	Adp33617 Mouse mtc
33	14.6	81.1	888	3	AAZ57117	Aaz57117 Mouse sta
34	14.6	81.1	888	4	AAAF1040	Aaf1040 Murine et
35	14.6	81.1	921	4	ABLI07179	Abli07179 Drosophila
36	14.6	81.1	1005	6	ABN21414	Abn21414 Human ORF
37	14.6	81.1	1041	10	ABZ66670	Abz66670 Orthosomy
38	14.6	81.1	1068	8	ABZ37539	Abz37539 Streptomy
39	14.6	81.1	1092	13	ADT41741	Adt41741 Bacterial
40	14.6	81.1	1209	13	ADQ45831	Adq45831 Bacterial
41	14.6	81.1	1424	4	ABLI02237	Abli02237 Drosophila
42	14.6	81.1	1674	13	ADQ60159	Adq60159 Bacterial
43	14.6	81.1	1707	13	ADQ46460	Adq46460 Bacterial
44	14.6	81.1	1749	6	AAAD21928	Aad21928 M. rosari
45	14.6	81.1	2366	11	ADJ18905	Adj18905 Human NOV
46	14.6	81.1	2534	12	ADQ63227	Adq63227 Novel hum
47	14.6	81.1	2534	11	ADQ63261	Adq63261 Human CDN
48	14.6	81.1	2996	4	ABLI07178	Abli07178 Drosophila
49	14.6	81.1	3380	4	ABLI07174	Abli07174 Drosophila
50	14.6	81.1	3500	4	ABLI02236	Abli02236 Drosophila
51	14.6	81.1	3799	13	ADQ48391	Adq48391 Bacterial
52	14.6	81.1	4360	4	ABLI21760	Abli21760 Drosophila
53	14.6	81.1	4979	5	AAAS91509	Aas91509 DNA encod
54	14.6	81.1	4979	12	ADQ32882	Adq32882 Nucleotid
55	14.6	81.1	5874	10	ADG73125	Adg73125 DNA encod
56	14.6	81.1	5874	10	ADLI2181	Adli2181 Pseudom
57	14.6	81.1	5997	6	AAI68598	Aai68598 A. cryptu
58	14.6	81.1	6946	8	ADLI24581	Adli24581 Human mal
59	14.6	81.1	7564	10	ABZ24581	Abz24581 Human cel
60	14.6	81.1	7616	13	ACN42378	Acn42378 Human dia
61	14.6	81.1	7838	4	ABLI21408	Abli21408 Drosophila
62	14.6	81.1	8931	3	AAAF75307	Aaf75307 DNA sequ
63	14.6	81.1	11188	6	AAAD21912	Aad21912 Micromono
64	14.6	81.1	14061	6	ABV93363	Abv93363 Human NOV
65	14.6	81.1	14109	6	ABV93362	Abv93362 Human NOV
66	14.6	81.1	17083	10	ABZ58812	Abz58812 S. cinna
67	14.6	81.1	20489	12	ADQ22881	Adq22881 Human sof
68	14.6	81.1	24120	8	ABX11642	Abx11642 Human ser
69	14.6	81.1	43865	13	ABD33215	Abd33215 Human can
70	14.6	81.1	45055	10	ABZ66808	Abz66808 Orthosomy
71	14.6	81.1	59816	8	ABZ37516	Abz37516 Streptomy
72	14.6	81.1	59816	8	ABZ37515	Abz37515 Streptomy
73	14.6	81.1	71292	11	ACN45142	Acn45142 Human gen
74	14.6	81.1	110000	4	AAI99682_34	Continuation (35 o
75	14.6	81.1	110000	4	AAI99683_34	Continuation (35 o
76	14.6	81.1	349980	6	ABO81847	Abog81847 Bifidobac
77	14.6	81.1	349980	6	ABO81848	Abog81848 Bifidobac
78	14.6	80.0	39	6	ABN84675	Abn84675 Vascular
79	14.4	80.0	144	6	ABK3123	Abk3123 Human pro
80	14.4	80.0	455	9	ACH15097	Ach15097 Human adu
81	14.4	80.0	604	8	ABX98613	Abx98613 Rice leaf
82	14.4	80.0	1116	13	ADT41618	Adt41618 Bacterial
83	14.4	80.0	1116	13	ADQ64029	Adq64029 Bacterial
84	14.4	80.0	1136	13	ADQ63654	Adq63654 Bacterial
85	14.4	80.0	1185	13	ADT48168	Adt48168 Bacterial
86	14.4	80.0	1644	13	ACN42143	Acn42143 Human dia
87	14.4	80.0	1981	11	ACN42142	Acn42142 Human dia
88	14.4	80.0	110000	11	ACN43998_1	Continuation (2 of
89	14.4	80.0	110000	11	ACN43998_1	Continuation (2 of
90	14.4	80.0	121062	12	ADQ97313	Adq97313 Human can
91	14.2	78.9	471	6	AAAD35503	Aad35503 Methylomo
92	14.2	78.9	471	6	ABK83262	Abk83262 High grow
93	14.2	78.9	471	6	ABK50085	Methylomo

C 94	14.2	78.9	579	11	ADJ12065	Adj12065 Maize cDN	167	14	77.8	783	11	ACH94524	Ach94524 Klebsiell
C 95	14.2	78.9	593	12	ADQ92250	Adq92250 Human aut	168	14	77.8	789	13	AAI97484	AAI97484 Human neu
C 96	14.2	78.9	796	2	AAV08844	AAV08844 Gene No.	169	14	77.8	789	13	ADSB3110	ADSB3110 Bacterial
C 97	14.2	78.9	897	13	ADSB3777	ADSB3777 Bacterial	170	14	77.8	789	13	ADSB2648	ADSB2648 Bacterial
C 98	14.2	78.9	897	13	ADSB64161	ADSB64161 Bacterial	171	14	77.8	789	13	ADSB59467	ADSB59467 Bacterial
C 99	14.2	78.9	909	13	ADSB63409	ADSB63409 Bacterial	172	14	77.8	840	12	ADNB1501	ADNB1501 Human KPP
C 100	14.2	78.9	1295	3	AAZ40192	AAZ40192 H. vulgar	173	14	77.8	840	13	ADSB59901	ADSB59901 Bacterial
C 101	14.2	78.9	1395	3	AAH78786	AAH78786 Nucleotid	C 174	14	77.8	874	4	AAK93113	AAK93113 Human cDN
C 102	14.2	78.9	1832	6	AAAL45918	AAAL45918 Terrabact	C 175	14	77.8	874	4	AAK91598	AAK91598 Human cDN
C 103	14.2	78.9	1836	13	ADSB64109	ADSB64109 Bacterial	C 176	14	77.8	874	12	ADL29540	ADL29540 5' end of
C 104	14.2	78.9	1944	11	ACH95422	ACH95422 Klebsiell	C 177	14	77.8	874	12	ADL28025	ADL28025 5' end of
C 105	14.2	78.9	2035	13	ADSB63719	ADSB63719 Bacterial	C 178	14	77.8	878	4	AAH03966	AAH03966 Human cDN
C 106	14.2	78.9	2035	13	ADSB63352	ADSB63352 Bacterial	C 179	14	77.8	885	11	ADBI17960	ADBI17960 Pseudomon
C 107	14.2	78.9	2376	16	ADSB62076	ADSB62076 Bacterial	C 180	14	77.8	900	13	ADBI1256	ADBI1256 Anti-biof
C 108	14.2	78.9	2627	6	ABSB69922	ABSB69922 Aspergill	C 181	14	77.8	918	11	ADMA4942	ADMA4942 Insect re
C 109	14.2	78.9	3425	12	ADQ63485	ADQ63485 Novel hum	C 182	14	77.8	956	13	ADRB8180	ADRB8180 Aspergill
C 110	14.2	78.9	3702	4	AAAS41970	AAAS41970 Genomic s	C 183	14	77.8	979	4	ABAA77110	ABAA77110 Profilera
C 111	14.2	78.9	3842	2	AAAT70153	AAAT70153 S. longisp	C 184	14	77.8	1041	10	ABZ66675	ABZ66675 Orthosomy
C 112	14.2	78.9	6085	2	AAAT70153	AAAT70153 S. longisp	C 185	14	77.8	1041	10	ABZ66674	ABZ66674 Orthosomy
C 113	14.2	78.9	9006	5	AAAT70153	AAAT70153 S. longisp	C 186	14	77.8	1047	8	ABZ37563	ABZ37563 Streptomy
C 114	14.2	78.9	34719	12	ADP90617	ADP90617 Micromono	C 187	14	77.8	1047	10	ABZ66673	ABZ66673 Orthosomy
C 115	14.2	78.9	110000	11	ADM27081_10	ADM27081_10 o	C 188	14	77.8	1047	10	ABZ66673	ABZ66673 Orthosomy
C 116	14.2	78.8	138	4	ADQ93471	ADQ93471 Human col	C 189	14	77.8	1089	8	ABZ37517	ABZ37517 Streptomy
C 117	14.2	78.8	139	4	ADQ93471	ADQ93471 Human col	C 190	14	77.8	1115	3	AAAC8229	AAAC8229 Arabidops
C 118	14.2	78.8	139	4	ADQ93471	ADQ93471 Human col	C 191	14	77.8	1124	10	ADK58797	ADK58797 plant DNA
C 119	14.2	78.8	139	4	ADQ93471	ADQ93471 Human col	C 192	14	77.8	1124	10	ADK58797	ADK58797 plant DNA
C 120	14.2	78.8	229	10	AAAL1628	AAAL1628 Human sec	C 193	14	77.8	1167	13	ADT18637	ADT18637 Bacterial
C 121	14.2	78.8	229	10	AAAL1628	AAAL1628 Human sec	C 194	14	77.8	1167	2	AAV84507	AAV84507 Human sec
C 122	14.2	78.8	255	8	AAAT24619	AAAT24619 Human gen	C 195	14	77.8	1167	2	AAV84507	AAV84507 Human sec
C 123	14.2	78.8	268	4	AAAT29169	AAAT29169 Streptomy	C 196	14	77.8	1167	9	ABAB83290	ABAB83290 Human sec
C 124	14.2	78.8	268	4	AAAT29169	AAAT29169 Streptomy	C 197	14	77.8	1167	9	ACH04791	ACH04791 Novel hum
C 125	14.2	78.8	291	2	AAAX8111	AAAX8111 Nucleotid	C 198	14	77.8	1167	10	ACD44601	ACD44601 Human cDN
C 126	14.2	78.8	303	2	ACA31973	ACA31973 Prokaryot	C 199	14	77.8	1179	4	AAAS07129	AAAS07129 DNA encod
C 127	14.2	78.8	323	2	AAAX97665	AAAX97665 Extended	C 200	14	77.8	1182	13	ADSA5817	ADSA5817 Bacterial
C 128	14.2	78.8	323	2	ADP18932	ADP18932 Human sec	C 201	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 129	14.2	78.8	335	2	AAAX1943	AAAX1943 Human sec	C 202	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 130	14.2	78.8	335	2	AAAX1943	AAAX1943 Human sec	C 203	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 131	14.2	78.8	348	4	AAAS37034	AAAS37034 Novel hum	C 204	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 132	14.2	78.8	354	8	ABE271757	ABE271757 Human can	C 205	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 133	14.2	78.8	356	6	ABN76664	ABN76664 Human ORF	C 206	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 134	14.2	78.8	411	2	AAAX1677	AAAX1677 Human sec	C 207	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 135	14.2	78.8	413	9	ACH16964	ACH16964 Human sec	C 208	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 136	14.2	78.8	418	9	ACH29501	ACH29501 Human adu	C 209	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 137	14.2	78.8	451	8	ABX36177	ABX36177 Bovine ES	C 210	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 138	14.2	78.8	464	9	ADBA46101	ADBA46101 Rice thia	C 211	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 139	14.2	78.8	466	10	ADCT5120	ADCT5120 Rice phyt	C 212	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 140	14.2	78.8	467	9	ACH43762	ACH43762 Human foe	C 213	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 141	14.2	78.8	469	9	ACH45256	ACH45256 Human foe	C 214	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 142	14.2	78.8	477	8	ACH19112	ACH19112 Prokaryot	C 215	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 143	14.2	78.8	490	2	AAAX1313	AAAX1313 Human sec	C 216	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 144	14.2	78.8	495	5	ABAI12903	ABAI12903 Human ner	C 217	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 145	14.2	78.8	540	10	ADDE6966	ADDE6966 Human lun	C 218	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 146	14.2	78.8	540	10	ADDE88220	ADDE88220 Human lun	C 219	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 147	14.2	78.8	570	11	ABD03664	ABD03664 Pseudomon	C 220	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 148	14.2	78.8	601	3	AAAC3541	AAAC3541 Arabidops	C 221	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 149	14.2	78.8	613	3	ABAI18510	ABAI18510 Human ner	C 222	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 150	14.2	78.8	618	6	ABQ66231	ABQ66231 Arabidops	C 223	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 151	14.2	78.8	621	12	ADN97658	ADN97658 S ambofac	C 224	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 152	14.2	78.8	636	9	ADBA6121	ADBA6121 Rice thia	C 225	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 153	14.2	78.8	640	10	ADCT6353	ADCT6353 DNA homol	C 226	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 154	14.2	78.8	644	10	ADCT6353	ADCT6353 DNA homol	C 227	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 155	14.2	78.8	648	11	ABD03882	ABD03882 Pseudomon	C 228	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 156	14.2	78.8	660	10	ADCT6356	ADCT6356 DNA homol	C 229	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 157	14.2	78.8	666	6	ABO65868	ABO65868 Arabidops	C 230	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 158	14.2	78.8	671	6	ADN09605	ADN09605 HIV-1 cod	C 231	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 159	14.2	78.8	671	6	ADN09605	ADN09605 HIV-1 cod	C 232	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 160	14.2	78.8	671	6	ADN09605	ADN09605 HIV-1 cod	C 233	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 161	14.2	78.8	671	6	ADN09605	ADN09605 HIV-1 cod	C 234	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 162	14.2	78.8	728	3	AAAC48877	AAAC48877 Arabidops	C 235	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 163	14.2	78.8	731	3	AAAC4504	AAAC4504 Arabidops	C 236	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 164	14.2	78.8	750	11	ABD06013	ABD06013 Pseudomon	C 237	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 165	14.2	78.8	763	10	ADCT6834	ADCT6834 Human adi	C 238	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon
C 166	14.2	78.8	774	10	ACF67152	ACF67152 Photornab	C 239	14	77.8	1251	11	ABDI17850	ABDI17850 Pseudomon



C 240	14	77.8	1524	6	ABN89320	Abn89320 Salmone1
C 241	14	77.8	1548	4	AAK94131	AAK94131 Human ful
C 242	14	77.8	1548	12	ADL30597	ADL30597 Full leng
C 243	14	77.8	1559	8	ACA31456	ACA31456 Prokaryot
C 244	14	77.8	1563	4	AAH17993	AAH17993 Human CDN
C 245	14	77.8	1593	8	ACA37741	ACA37741 Prokaryot
C 246	14	77.8	1647	11	ABD06454	ABD06454 Pseudomon
C 247	14	77.8	1660	6	ABK83216	ABK83216 Human tra
C 248	14	77.8	1680	6	ABK75035	ABK75035 Bacillus
C 249	14	77.8	1701	11	ADP75211	ADP75211 Human ADA
C 250	14	77.8	1811	6	AA662830	AA662830 cDNA sequ
C 251	14	77.8	1857	8	ACA27195	ACA27195 Prokaryot
C 252	14	77.8	1863	8	ABZ72016	ABZ72016 Human cha
C 253	14	77.8	1863	13	ACN37721	ACN37721 Tumour-as
C 254	14	77.8	1863	13	ADP54117	ADP54117 Human PRO
C 255	14	77.8	1866	8	ABT32135	ABT32135 Benzodiaz
C 256	14	77.8	1867	10	ACA92442	ACA92442 DNA encod
C 257	14	77.8	1879	11	AD131416	AD131416 Human CDN
C 258	14	77.8	1882	11	ADM02350	ADM02350 Human CDN
C 259	14	77.8	1908	2	AAQ99364	AAQ99364 S. livida
C 260	14	77.8	1908	2	AAV84065	AAV84065 Tripeptid
C 261	14	77.8	1908	3	AAAC61403	AAAC61403 cDNA sequ
C 262	14	77.8	1920	2	AAZ06460	AAZ06460 Nucleotid
C 263	14	77.8	1920	4	ABL26517	ABL26517 Drosophil
C 264	14	77.8	1980	11	ABD05586	ABD05586 Pseudomon
C 265	14	77.8	2023	2	AAV06593	AAV06593 Human N-P
C 266	14	77.8	2023	10	ABSS7455	ABSS7455 Human sho
C 267	14	77.8	2070	11	ABD03585	ABD03585 Pseudomon
C 268	14	77.8	2100	8	ADA50514	ADA50514 Human pro
C 269	14	77.8	2103	8	ACA37871	ACA37871 Prokaryot
C 270	14	77.8	2166	12	AD005545	AD005545 Bradythiz
C 271	14	77.8	2175	6	ABA99462	ABA99462 Actinopia
C 272	14	77.8	2176	6	ABL90432	ABL90432 Human pol
C 273	14	77.8	2179	4	AAH17168	AAH17168 Human CDN
C 274	14	77.8	2334	12	ADU40133	ADU40133 Plant CDN
C 275	14	77.8	2394	13	ADS48448	ADS48448 Bacterial
C 276	14	77.8	2426	11	ADM02149	ADM02149 Human CDN
C 277	14	77.8	2427	3	AACT77157	AACT77157 Human ORF
C 278	14	77.8	2540	2	AAO51488	AAO51488 Ornithine
C 279	14	77.8	2580	12	ADJ94704	ADJ94704 Human euk
C 280	14	77.8	2582	2	AAT73117	AAT73117 Actinopia
C 281	14	77.8	2672	4	ABL06124	ABL06124 Drosophil
C 282	14	77.8	2712	8	ACA26785	ACA26785 Prokaryot
C 283	14	77.8	2817	8	ACA41660	ACA41660 Prokaryot
C 284	14	77.8	2827	4	AAAF31253	AAAF31253 Human imm
C 285	14	77.8	2956	13	ADR84593	ADR84593 Aspergill
C 286	14	77.8	2958	5	AA929666	AA929666 DNA encod
C 287	14	77.8	2960	5	AA990160	AA990160 DNA encod
C 288	14	77.8	2983	4	AAH14446	AAH14446 Human CDN
C 289	14	77.8	3011	6	ABO99304	ABO99304 Human cod
C 290	14	77.8	3018	8	AA161203	AA161203 Actinomy
C 291	14	77.8	3103	12	ADQ09942	ADQ09942 Mouse NK-
C 292	14	77.8	3103	12	ADQ09868	ADQ09868 Mouse NK-
C 293	14	77.8	3110	12	ADQ24300	ADQ24300 Human sof
C 294	14	77.8	3144	4	AAH47048	AAH47048 Larynx ca
C 295	14	77.8	3216	12	AD134850	AD134850 Bovine pr
C 296	14	77.8	3288	4	ABL06132	ABL06132 Drosophil
C 297	14	77.8	3561	6	AA997178	AA997178 Human met
C 298	14	77.8	3567	6	ABK12893	ABK12893 Human pro
C 299	14	77.8	3580	6	AB878717	AB878717 Human CDN
C 300	14	77.8	3624	9	ACC58643	ACC58643 Human ADA

## ALIGNMENTS

RESULT 1  
ID ABN86366 standard; DNA; 18 BP.  
AC ABN86366;  
XX  
XX

DT 21-OCT-2002 (first entry)

XX	XX	S. coelicolor scdA gene internal segment amplifying primer 1.
DE	XX	Antibiotic; bacterium; scdA; afcA; scdR; arpA; barA; actinorhodine; Act;
KM	XX	undecylprodigiosin; Red; PCR; primer; ss.
KW	XX	Streptomycetes coelicolor.
OS	XX	CA2322241-A1.
PN	XX	23-APR-2002.
PD	XX	23-OCT-2000; 2000CA-02322241.
PF	XX	23-OCT-2000; 2000CA-02322241.
PR	XX	23-OCT-2000; 2000CA-02322241.
XX	XX	(PLAN-) PLANT BIOSCIENCE LTD.
PA	XX	Takano E, Bibb M;
PI	XX	WPI; 2002-501089/54.
DR	XX	Modifying antibiotic-producing Streptomycetes, to increase, or alter timing
XX	XX	of, antibiotic production, by deleting the scdA or scdR genes.
PT	XX	Claim 19; Page 53; 64pp; English.
XX	XX	The invention provides a method for modifying an antibiotic-producing
PS	XX	strain of Streptomycetes to increase production of antibiotics or to alter
CC	XX	the timing of antibiotic production. The modification is functional
CC	XX	deletion of the scdA gene of S. coelicolor, or its homologues, but is not
CC	XX	functional deletion of the afcA gene of S. griseus, or the modification is
CC	XX	functional deletion of the scdR gene of S. coelicolor, or its homologues,
CC	XX	but is not deletion of arpA of S. griseus nor barA of S. virginiae. The
CC	XX	method is particularly used for production of the antibiotics
CC	XX	actinorhodine (Act) and undecylprodigiosin (Red). The present sequence
CC	XX	represents a PCR primer for amplifying an internal segment of the scdA
CC	XX	gene from S. coelicolor M145 total DNA
SO	XX	Sequence 18 BP; 3 A; 6 C; 5 G; 2 T; 0 U; 2 Other;
QY	XX	Query Match 95.6%; Score 17.2; DB 6; Length 18;
DB	XX	Best Local Similarity 100.0%; Pred. No. 1.5e+02;
DB	XX	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB	XX	1 GACCACTGCTCCGGCATTG 18
DB	XX	1 GACCACTGCTCCGGCATTG 18
RESULT 2	XX	ADBI4795
ID	XX	ADBI4795 standard; DNA; 18 BP.
AC	XX	ADBI4795;
XX	XX	29-JAN-2004 (first entry)
DE	XX	PCR primer 1 related to Streptomycetes coelicolor antibiotic production.
KM	XX	*antibiotic-producing strain; antibiotic production; scdA gene; scdR gene;
KW	XX	actinorhodin; undecylprodigiosin; PCR; primer; ss.
OS	XX	Synthetic.
OS	XX	Streptomycetes coelicolor.
PN	XX	US2003124644-A1.
PD	XX	03-JUL-2003.
PF	XX	23-OCT-2001; 2001US-00017471.
PR	XX	23-OCT-2000; 2000US-0242561P.

XX (TAKA/) TAKANO E.  
PA (BIBB/) BIBB M J.  
XX  
PI Takano E, Bibb MJ;  
XX WPI; 2003-810983/76.  
DR  
XX Modifying an antibiotic-producing strain of Streptomyces coelicolor or  
PT Streptomyces lividans to increase or alter the timing of antibiotic  
PR production in the strain, comprises functionally deleting in the strain  
PT the scdA or scdR gene.  
XX  
PS Example 8; Page 9; 33pp; English.  
XX  
XX This invention relates to the novel modification of an antibiotic-  
CC producing strain of Streptomyces coelicolor or Streptomyces lividans to  
CC increase or to alter the timing of antibiotic production in the strain.  
CC The method comprises functionally deleting in the strain the scdA or scdR  
CC gene. The method is useful in increasing and altering the timing of  
CC antibiotic production (especially actinorhodin and undecylprodigiosin) in  
CC Streptomyces species, particularly Streptomyces coelicolor or  
CC Streptomyces lividans. The present sequence is that of a degenerate PCR  
CC primer which was used for amplification of a region of the Streptomyces  
CC coelicolor scdA gene during the exemplification of the invention.  
CC  
SQ Sequence 18 BP; 3 A; 6 C; 5 G; 2 T; 0 U; 2 Other;  
XX  
Query Match 95.6%; Score 17.2; DB 10; Length 18;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 16; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 GACCACTGTCGCGCATG 18  
|||  
1 GACCACTGTCGCGCATG 18  
DB  
XX  
RESULT 3  
ADQ48866  
ID ADQ48866 standard; DNA; 587 BP.  
XX  
AC ADQ48866;  
XX  
DT 21-OCT-2004 (first entry)  
XX  
XX Novel canine microarray-related DNA sequence SegID168.  
XX  
XX canine microarray; drug screening; toxicity assay;  
XX environmental pollutant; cellular response; gene expression profile;  
XX toxic response; liver necrosis; fatty liver disease;  
XX protein adduct formation; hepatitis; dog; ds.  
XX  
OS Canis familiaris.  
XX  
XX WO2004063324-A2.  
XX  
XX 29-JUL-2004.  
XX  
XX 05-MAY-2003; 2003WO-US013853.  
XX  
XX 03-MAY-2002; 2002US-0377240P.  
XX  
XX (GENE-) GENE LOGIC INC.  
XX (PFIZ) PFIZER PROD INC.  
XX  
XX Digians JC, Porter M, Wei T;  
XX  
XX WPI; 2004-561890/54.  
XX  
XX New isolated nucleic acid molecule, useful for drug screening and  
XX toxicity assay or for assessing the impact, including toxicity, of a  
XX compound, pharmaceutical agent or environmental pollutant on a cell or  
XX living organism.

XX  
PS Claim 1; SEQ ID NO 168; 41pp; English.  
XX  
XX This invention is related to a novel isolated canine nucleic acid  
CC sequences and the construction of canine microarrays containing a  
CC significant portion of the canine genome. The isolated canine nucleic  
CC acid sequences of the invention may be useful for drug screening and  
CC toxicity assays. The invention is therefore useful for assessing the  
CC impact, including toxicity, of a compound, pharmaceutical agent or  
CC environmental pollutant on a cell or living organism. The methods are  
CC useful for detecting genes that are up- or down-regulated in canines in a  
CC disease state. The sequences are useful as diagnostic agents or markers  
CC to detect a cellular response in a sample individually or as part of a  
CC gene expression profile. It is also useful as a target for agents that  
CC modulate gene expression or activity. The database is useful for  
CC producing electronic Northern blots that allow the user to determine the cell  
CC type or tissue in which a given gene is expressed and to allow  
CC determination of the abundance or expression level of a given gene in a  
CC particular tissue or cell. The methods are useful for determining the  
CC similarity of a toxic response to one or more individual compounds. The  
CC methods are useful for predicting at least one toxic response or the  
CC likelihood that a compound or test agent will induce various specific  
CC pathologies such as those of the liver (liver necrosis, fatty liver  
CC disease, protein adduct formation or hepatitis), those of the kidney,  
CC heart, brain or testes, or other pathologies associated with at least one  
CC of the toxins. The methods are also useful for predicting or elucidating  
CC the potential cellular pathways influenced, induced or modulated by the  
CC compound or test agent due to the similarity of the expression profile  
CC compared to the profile induced by a known toxin. The present sequence is  
CC that of a canine DNA sequence which was claimed for use during the  
CC production of a canine microarray of the invention.  
XX  
SQ Sequence 587 BP; 81 A; 186 C; 195 G; 120 T; 0 U; 5 Other;  
XX  
Query Match 90.0%; Score 16.2; DB 13; Length 587;  
Best Local Similarity 88.2%; Pred. No. 4.1e+02;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 GACCACTGTCGCGCAT 17  
|||  
401 GACCACTGTCGCGCAT 417  
DB  
XX  
RESULT 4  
ABN86379/C  
ID ABN86379 standard; DNA; 4346 BP.  
XX  
XX ABN86379;  
XX  
XX 21-OCT-2002 (first entry)  
XX  
XX scdA, scdR and scdB encoding genes containing DNA sequence.  
XX  
XX Antibiotic; bacterium; scdA; afsA; scdR; arpA; barA; actinorhodine; Act;  
XX undecylprodigiosin; Red; gene; ds.  
XX  
XX Streptomyces coelicolor.  
XX  
XX CA2322241-A1.  
XX  
XX 23-APR-2002.  
XX  
XX 23-OCT-2000; 2000CA-02322241.  
XX  
XX 23-OCT-2000; 2000CA-02322241.  
XX  
XX (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
XX Takano E, Bibb M;  
XX  
XX WPI; 2002-501089/54.  
XX P-PSDB; ABB80940, ABB80941, ABB80942.  
XX

PT		Streptomycetes lividans to increase or alter the timing of antibiotic production in the strain, comprises functionally deleting in the strain the scdA or scdr gene.
PS	Claim 19; Fig 14; 33pp; English.	
XX		
CC	This invention relates to the novel modification of an antibiotic-	
CC	producing strain of Streptomycetes coelicolor or Streptomycetes lividans to	
CC	increase or to alter the timing of antibiotic production in the strain.	
CC	The method comprises functionally deleting in the strain the scdA or scdr	
CC	gene. The method is useful in increasing and altering the timing of	
CC	antibiotic production (especially actinorhodin and undecylprodigiosin) in	
CC	Streptomycetes species, particularly Streptomycetes coelicolor or	
CC	Streptomycetes lividans. The present sequence is that of a scdb region of the	
CC	Streptomycetes coelicolor genome, which encodes the scdA, scdb and scdr	
CC	proteins and which is related to the invention.	
XX		
SQ	Sequence 4346 BP; 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;	
	Query Match	88.9%; Score 16; DB 10; Length 4346;
	Best Local Similarity	88.9%; Pred. No. 4,7e+02;
	Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	
OY	1 GACCACGTCSCCGGCATG 18      ::	
DB	1453 GACCACGTACCGGCATG 1436	
RESULT 6		
AAC75505		
ID	AAC75505 standard; cDNA; 465 BP.	
XX		
AC	AAC75505;	
XX		
DT	08-FEB-2001 (first entry)	
DE		
XX	Human OREFX ORF1060 polynucleotide sequence SEQ ID NO:2119.	
XX		
KW	Human; open reading frame; OREFX; detection; cytostatic; hepatotropic;	
KV	nervary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;	
KW	anticancer; osteopathic; antiarthritic; immunosuppressant; cardiant;	
KM	immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;	
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;	
KV	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;	
KW	antihaemic; gene therapy; cancer; proliferative disorder; hypertension;	
KV	neurodegenerative disorder; osteoarthritis; graft vs host disease;	
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	
KV	cholesterol ester storage; systemic lupus erythematosus; infection;	
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	
KV	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;	
KW	bone damage; cattleage damage; antiinflammatory disease; coagulation;	
KV	thrombosis; contraceptive; ss.	
XX		
XS	Homo sapiens.	
XX		
PN	WO200058473-A2.	
XX		
PB	05-OCT-2000.	
XX		
PF	31-MAR-2000; 2000MO-USO08621.	
XX		
PR	31-MAR-1999; 99US-0127607P.	
XX	02-APR-1999; 99US-0127636P.	
PR	05-APR-1999; 99US-0127728P.	
XX	30-MAR-2000; 2000US-00540763.	
PA	(CURAGEN CORP.	
PI	Shimkets RA, Leach M;	
XX	WPI; 2000-602362/57.	
RR	P-PSDB; AAB41296.	
XX		

CC sequence P. aeruginosa is an opportunistic human pathogen present in  
CC soil water and plants. The specification describes virulence polypeptides  
CC and nucleic acid sequence encoding such polypeptides. These sequences can  
CC be used to identify a compound which is capable of decreasing the  
CC expression of a pathogenic virulence factor. Compounds that inhibit the  
CC expression or activity of virulence factor polypeptides can be used to  
CC treat pathogenic infections, especially where the infection is a P.  
CC aeruginosa infection. note: the sequences given in the specification were  
CC poorly legible, and in some instances assumptions were made as to the  
CC identity of the base; it is therefore possible that the sequence given  
CC below is not entirely correct

SQ Sequence 513 BP; 91 A; 174 C; 172 G; 76 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 2; Length 513;  
Best Local Similarity 83.3%; Pred. No. 7.9e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GACCACGTCGCCGCATG 18  
||| ||| : | : ||| ||  
Db 133 GACCACGTCTCCGGCGCATG 150

RESULT 8  
ADR91782/c  
ID ADR91782 standard; DNA; 612 BP.

XX ADR91782;  
XX AC  
XX AC  
DT 16-DEC-2004 (first entry)  
XX  
XX  
DE Novel S. pneumoniae DNA sequence, SEQ ID 417.

XX Meningitis; bacteraemia; pneumonia; otitis media; ds;  
XX bacterial infection.  
XX  
OS Streptococcus pneumoniae.

XX  
XX US6800744-B1.  
XX  
XX  
PD 05-OCT-2004.  
XX  
XX  
PF 30-JUN-1998; 98US-00107433.  
XX  
XX  
PR 02-JUL-1997; 97US-0051553P.  
XX PR 12-MAY-1998; 98US-0085131P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;  
DR WPI; 2004-697205/68.  
DR P-PSDB; ADR94385.  
XX  
XX  
PT New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX  
XX  
PS Disclosure; SEQ ID NO 417; 151bp; English.

CC The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR9136polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94409, ADR94480, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91866, ADR92197, ADR92234, ADR93039, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequence, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridisable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe

CC comprising at least 20 consecutive nucleotides of the nucleotide  
 CC sequences as cited above. The methods and compositions of the present  
 CC invention are useful for the diagnosis, prevention and/or treatment of  
 CC pathological conditions resulting from bacterial infection by  
 CC *Streptococcus pneumoniae* e.g. pneumonia, bacteraemia, meningitis and  
 CC otitis media. The present sequence is one of the 2603 disclosed S.  
 CC *pneumoniae* nucleic acid sequences. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from USPTO at  
 CC [seqdata.uspto.gov/sequence.html?docID=680074481](http://seqdata.uspto.gov/sequence.html?docID=680074481).

XX Sequence 612 BP; 164 A; 122 C; 135 G; 191 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 13; Length 612;

Best Local Similarity 83.3%; Pred. No. 7.8e+02;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGCTCCGGCATTG 18  
 |||||  
 Db 160 GACCACTGCTCCGGCATTG 143

# RESULT 9

ID ABL10881 standard; cDNA; 744 BP.

XX ABL10881;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 27125.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX MO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009221.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB66778.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX genes from *Drosophila* and for elucidating cell signalling and cell-cell  
 XX interactions.

XX Claim 1; SEQ ID NO 27125; 21pp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABH72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)

XX Sequence 744 BP; 174 A; 249 C; 192 G; 129 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 4; Length 744;

Best Local Similarity 83.3%; Pred. No. 7.8e+02;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGCTCCGGCATTG 18  
 |||||  
 Db 454 GTCCACGTGCTCCGGCATTG 471

# RESULT 10

ID AAX98245/c  
 AAX98245 standard; DNA; 766 BP.

XX AAX98245;

XX 25-OCT-1999 (first entry)

XX Nucleotide sequence of ORF4 of contig 1344.

XX Human pathogen; virulence polypeptide; virulence factor;

XX pathogenic infection; *Pseudomonas aeruginosa* infection; ss.

XX *Pseudomonas aeruginosa*.

XX Key Location/Qualifiers

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

/\*tag= a  
 /note= "these bases represent nucleotides missing from  
 the sequence given in the specification; They are  
 included to maintain the nucleotide numbering given in  
 the specification for this sequence"  
 641..650  
 /\*tag= b  
 /note= "these bases represent nucleotides missing from  
 the sequence given in the specification; They are  
 included to maintain the nucleotide numbering given in  
 the specification for this sequence"  
 681..700  
 /\*tag= c  
 /note= "these bases represent nucleotides missing from  
 the sequence given in the specification; They are  
 included to maintain the nucleotide numbering given in  
 the specification for this sequence"  
 721..730  
 /\*tag= d  
 /note= "these bases represent nucleotides missing from  
 the sequence given in the specification; They are  
 included to maintain the nucleotide numbering given in  
 the specification for this sequence"  
 761..766  
 /\*tag= e  
 /note= "these bases represent nucleotides missing from  
 the sequence given in the specification; They are  
 included to maintain the nucleotide numbering given in  
 the specification for this sequence"

MO9927129-A1.

03-JUN-1999.

25-NOV-1998; 98WO-US025247.

25-NOV-1997; 97US-0066517P.

(GENO ) GEN HOSPITAL CORP.

Ausubel F, Goodman HW, Rahne LG, Mahajan-Miklos S, Tan M, Cao H;

Drenkard E, Tsongalis J;

WPI; 1999-357851/30.

Virulence factors useful in developing disease treatments.

Disclosure; Fig 32; 228pp; English.

of *Pseudomonas* species using biochip technology. Sequences ABD01397-CC CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
The sequence data for this patent did not form part of the printed  
specification but was obtained in electronic format from USPTO at  
segdata.uspto.gov/sequence.html  
SQ Sequence 891 BP; 130 A; 286 C; 308 G; 167 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 11; Length 891;  
Best Local Similarity 83.3%; Pred. No. 7.7e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0

QY 1 GACCACTGTCSCGGCATG 18  
|||||:||||||  
DB 432 GACCAGTCCGCCGCATG 415

RESULT 12  
ABX05949/c  
ID ABX05949 standard; DNA, 1188 BP.  
XX  
XX AC ABX05949;  
XX  
DT 27-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
XX DE S. pneumoniae type 4 strain coding region #337.  
XX  
XX KM Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;  
KM ear infection; antiinflammatory; antibacterial; immunostimulant;  
KM auditory; respiratory; gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
PN WO200277021-A2.  
PD 03-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-IB002163.  
XX  
XX 27-MAR-2001; 2001GB-00007658.  
PR  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Maesmani V, Tettelin H, Frazer C;  
XX  
XX WPI. 2003-040579/03.  
DR P-PSDB; ABU00670.  
XX  
XX New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.

Claim 6; SEQ ID NO 473; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as ABS65444. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are a nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a *Streptococcus* nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified assay comprising contacting a test compound with the

CC protein, and determining whether the test compound binds to the protein  
CC and a Streptococcus pneumoniae bacterium, where one or more genes  
CC encoding the proteins has been rendered inactive. The proteins, nucleic  
CC acid molecules, antibody and compositions are useful as medicaments for  
CC treating or preventing a disease or infection due to streptococcus  
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
CC media or ear infection. They are also useful in developing vaccines,  
CC diagnostics and antibiotics. The methods are useful for identifying  
CC immunodominant proteins. The present sequence is one of the 2489  
CC identified coding region from the genomic sequence. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 27-OCT-2003 to  
CC standardise OS field)

XX  
SQ Sequence 1188 BP; 355 A; 228 C; 261 G; 344 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 10; Length 1188;  
Best Local Similarity 83.3%; Pred. No. 7.7e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGCATG 18  
|||||:|:|:|  
Db 739 GACCACTGTCGCGCATG 722

RESULT 13  
ABD15034/c  
ID ABD15034 standard; DNA; 1275 BP.

XX  
AC ABD15034;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polynucleotide #13638.  
XX  
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
XX antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX WPI; 2003-615309/58.  
XX P-PSDB; AB081463.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 13638; 455pp; English.

CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused

CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biotin technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html

XX  
SQ Sequence 1275 BP; 208 A; 409 C; 443 G; 215 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 11; Length 1275;  
Best Local Similarity 83.3%; Pred. No. 7.6e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGCATG 18  
|||||:|:|:|  
Db 367 GACCACTGTCGCGCATG 350

RESULT 14  
AAZ32291  
ID AAZ32291 standard; DNA; 1408 BP.

XX  
AC AAZ32291;  
XX  
DT 20-JAN-2000 (first entry)  
XX  
DE Sphingomonas sp. biotin synthase biob nucleotide sequence.  
XX  
KW Sphingomonas; biotin; biotin synthase; deethiobiotin; fermentation; ds.  
XX  
OS Sphingomonas sp.  
XX  
PN JP11276175-A.  
XX  
PD 12-OCT-1999.  
XX  
PF 31-MAR-1998; 98JP-00086972.  
XX  
PR 31-MAR-1998; 98JP-00086972.  
XX  
PA (SUMO) SUMITOMO CHEM CO LTD.  
XX  
DR WPI; 1999-626939/54.  
XX  
PT Preparation of biotin - comprising fermentation using Escherichia coli.  
XX  
PS Example; Page 11-12; 13pp; Japanese.  
XX  
CC A method has been developed for the preparation of biotin by fermentation  
CC using a microbe in which Escherichia coli having a plasmid containing a  
CC gene coding an enzyme having biotin synthase activity is added to the  
CC culture liquid. The method can convert deethiobiotin accumulated in the  
CC culture liquid to biotin efficiently. The present sequence represents a  
CC Sphingomonas sp. gene having biotin synthase activity, used in the  
CC exemplification of the present invention

XX  
SQ Sequence 1408 BP; 214 A; 540 C; 442 G; 212 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 2; Length 1408;  
Best Local Similarity 83.3%; Pred. No. 7.6e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGCATG 18  
|||||:|:|:|  
Db 1057 GACCACTGTCGCGCATG 1074

RESULT 15  
AAZ32287  
ID AAZ32287 standard; DNA; 1408 BP.

XX  
AC AAZ32287;  
XX

```
DT 20-JAN-2000 (first entry)
XX Sphingomonas sp. biotin synthase biof nucleotide sequence.
DE Sphingomonas; biotin; biotin synthase; desbiobiotin; fermentation; ds.
XX Sphingomonas sp.
OS Sphingomonas sp.
XX JP11276175-A.
PN 12-OCT-1999.
PD 31-MAR-1998; 98JP-00086972.
XX 31-MAR-1998; 98JP-00086972.
PR 31-MAR-1998; 98JP-00086972.
XX (SUMO ) SUMITOMO CHEM CO LTD.
PA WPI, 1999-626939/54.
XX WPI, 1999-626939/54.
DR WPI, 1999-626939/54.
XX Preparation of biotin - comprising fermentation using Escherichia coli.
PT Example; Page 9-10; 13pp; Japanese.
PS Example; Page 9-10; 13pp; Japanese.
XX A method has been developed for the preparation of biotin by fermentation
CC using a microbe in which Escherichia coli having a plasmid containing a
CC gene coding an enzyme having biotin synthase activity is added to the
CC culture liquid. The method can convert desbiobiotin accumulated in the
CC culture liquid to biotin efficiently. The present sequence represents a
CC Sphingomonas sp. gene having biotin synthase activity, used in the
CC exemplification of the present invention.
XX Sequence 1408 BP; 214 A; 541 C; 441 G; 212 T; 0 U; 0 Other;
SQ Query Match 86.7%; Score 15.6; DB 2; Length 1408;
Best Local Similarity 83.3%; Pred. No. 7.6e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY 1 GACCAAGTSCGCGCATG 18
DB 1057 GACCTCGTCCGCGCATG 1074
RESULT 16
ADA73004
XX ADA73004 standard; DNA; 2000 BP.
AC ADA73004;
XX 20-NOV-2003 (first entry)
DT 20-NOV-2003 (first entry)
XX Rice gene, SEQ ID 6330.
DE Rice gene, SEQ ID 6330.
XX Plant; bacterial infection; fungal infection; viral infection; rice;
KM gene; ds.
XX Oryza sativa.
OS Oryza sativa.
XX MO2003000898-A1.
PN MO2003000898-A1.
XX 03-JAN-2003.
PD 03-JAN-2003.
PF 22-JUN-2001; 2001WO-1B001105.
XX 22-JUN-2001; 2001WO-1B001105.
PR 22-JUN-2001; 2001WO-1B001105.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
PI WPI, 2003-175290/17.
DR WPI, 2003-175290/17.
XX
```

```
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX Claim 27; SEQ ID NO 6330; 899p; English.
PS Claim 27; SEQ ID NO 6330; 899p; English.
XX The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC the expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX Sequence 2000 BP; 419 A; 529 C; 502 G; 547 T; 0 U; 3 Other;
SQ Query Match 86.7%; Score 15.6; DB 8; Length 2000;
Best Local Similarity 83.3%; Pred. No. 7.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY 1 GACCAAGTSCGCGCATG 18
DB 1089 GGCAAGTCCCGCATG 1106
RESULT 17
AA98244
XX AA98244 standard; DNA; 2048 BP.
AC AA98244;
XX 25-OCT-1999 (first entry)
DT 25-OCT-1999 (first entry)
XX Contig 1344 identified using sequence tag 33C7.
DE Contig 1344 identified using sequence tag 33C7.
XX Human pathogen; virulence polypeptide; virulence factor;
KM pathogenic infection; Pseudomonas aeruginosa infection; ss.
XX Pseudomonas aeruginosa.
OS Pseudomonas aeruginosa.
XX Key
FH Location/Qualifiers
FT 1141..1444
FT misc_feature
FT /tag= a
FT /note= "these bases represent nucleotides missing from
FT the sequence given in the specification; They are
FT included to maintain the nucleotide numbering given in
FT the specification for this sequence"
FT 1201..1210
FT misc_feature
FT /tag= b
FT /note= "these bases represent nucleotides missing from
FT the sequence given in the specification; They are
FT included to maintain the nucleotide numbering given in
FT the specification for this sequence"
FT 1261..1290
FT misc_feature
FT /tag= c
FT /note= "these bases represent nucleotides missing from
FT the sequence given in the specification; They are
FT included to maintain the nucleotide numbering given in
FT the specification for this sequence"
FT 1321..1350
FT misc_feature
FT /tag= d
FT /note= "these bases represent nucleotides missing from
FT the sequence given in the specification; They are
FT included to maintain the nucleotide numbering given in
FT the specification for this sequence"
FT 1371..1400
FT misc_feature
FT /tag= e
FT /note= "these bases represent nucleotides missing from
FT the sequence given in the specification; They are
```



FT included to maintain the nucleotide numbering given in  
 FT the specification for this sequence"  
 FT 1431..1460  
 FT /\*tag= f  
 FT /note= "these bases represent nucleotides missing from  
 FT the sequence given in the specification; they are  
 FT included to maintain the nucleotide numbering given in  
 FT the specification for this sequence"  
 XX  
 XX WO9927129-A1.  
 XX  
 XX 03-JUN-1999.  
 XX  
 XX 25-NOV-1998; 98WO-US025247.  
 XX  
 XX 25-NOV-1997; 97US-0066517P.  
 XX  
 XX (GEHO ) GEN HOSPITAL CORP.  
 XX  
 XX Ausubel F, Goodman HW, Rahne LG, Mahajan-Miklos S, Tan M, Cao H,  
 XX Drenkard E, Tsongalis J;  
 XX WPI, 1999-357851/30.  
 XX  
 XX virulence factors useful in developing disease treatments.  
 XX  
 XX Disclosure; Fig 32; 228pp; English.  
 XX  
 XX The present sequence represents a Pseudomonas aeruginosa nucleic acid  
 XX sequence. P. aeruginosa is an opportunistic human pathogen present in  
 XX soil water and plants. The specification describes virulence polypeptides  
 XX and nucleic acid sequence encoding such polypeptides. These sequences can  
 XX be used to identify a compound which is capable of decreasing the  
 XX expression of a pathogenic virulence factor. Compounds that inhibit the  
 XX expression or activity of virulence factor polypeptides can be used to  
 XX treat pathogenic infections, especially where the infection is a P.  
 XX aeruginosa infection. note: the sequences given in the specification were  
 XX poorly legible, and in some instances assumptions were made as to the  
 XX identity of the bases; it is therefore possible that the sequence given  
 XX below is not entirely correct  
 XX  
 XX Sequence 2048 BP; 322 A; 676 C; 594 G; 322 T; 0 U; 134 Other;  
 XX  
 XX Query Match 86.7%; Score 15.6; DB 2; Length 2048;  
 XX Best Local Similarity 83.3%; Pred. No. 7.5e+02;  
 XX Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 XX 1 GACCACTGCTCCGGGCGATG 18  
 XX |||||:|:|:|:|:|:|  
 XX Db 1069 GACCACTGCTCCGGGCGATG 1086  
 XX  
 XX RESULT 18  
 XX AAV52338/c  
 XX ID AAV52338 standard; DNA; 2395 BP.  
 XX  
 XX AAV52338;  
 XX  
 XX 23-OCT-1998 (first entry)  
 XX  
 XX Streptococcus pneumoniae genome fragment SEQ ID NO:205.  
 XX  
 XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 XX computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX  
 XX Streptococcus pneumoniae.  
 XX  
 XX WO9818931-A2.  
 XX  
 XX 07-MAY-1998.  
 XX  
 XX 30-OCT-1997; 97WO-US019588.  
 XX

PR 31-OCT-1996; 96US-0029960P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;  
 XX Dougherty BA;  
 XX WPI, 1998-272225/24.  
 XX  
 XX Computer-readable medium with recorded Streptococcus pneumoniae  
 XX polynucleotide sequences - useful in diagnostic kits and assays, and  
 XX pharmaceutical compositions and vaccines for Streptococcus pneumoniae.  
 XX  
 XX Claim 1; Page 1189-1191, 1409pp; English.  
 XX  
 XX The present invention describes a computer readable medium which has the  
 XX nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded  
 XX on it, or a representative fragment or a sequence at least 95% identical  
 XX to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 XX to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus  
 XX pneumoniae. The present invention also describes an isolated nucleic acid  
 XX molecule encoding a homologue of any of the fragments of the S. pneumoniae  
 XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 XX by a process comprising: (a) screening a genomic DNA library using as a  
 XX probe a target sequence defined by any of the sequences in SEQ ID NO:1 to  
 XX 391, identifying members of the library which contain sequences that  
 XX hybridize to the target sequence and isolating the nucleic acid molecules  
 XX from the members; or (b) isolating mRNA, DNA or cDNA produced from an  
 XX organism, amplifying nucleic acid molecules whose nucleotide sequence is  
 XX homologous to amplification primers derived from the fragment of the S.  
 XX pneumoniae genome to prime the amplification and isolating the amplified  
 XX sequences. The computer readable medium can be used in a computer-based  
 XX system for identifying fragments of the S. pneumoniae genome of  
 XX commercial importance, or expression modulating fragments of the S.  
 XX pneumoniae genome. Products from the present invention can be used in  
 XX diagnosis kits and assays, and pharmaceutical compositions and vaccines  
 XX for S. pneumoniae  
 XX  
 XX Sequence 2395 BP; 755 A; 415 C; 548 G; 677 T; 0 U; 0 Other;  
 XX  
 XX Query Match 86.7%; Score 15.6; DB 2; Length 2395;  
 XX Best Local Similarity 83.3%; Pred. No. 7.5e+02;  
 XX Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 XX 1 GACCACTGCTCCGGGCGATG 18  
 XX |||||:|:|:|:|:|:|  
 XX Db 1257 GACCACTGCTCCGGGCGATG 1240  
 XX  
 XX RESULT 19  
 XX AEN59696  
 XX ID AEN59696 standard; cDNA; 5042 BP.  
 XX  
 XX AEN59696;  
 XX  
 XX 28-JUN-2002 (first entry)  
 XX  
 XX Novel human coding sequence SEQ ID NO: 107.  
 XX  
 XX Human; anti-infective; vaccine; anti-inflammatory; immunomodulator;  
 XX human; anti-infective; vaccine; anti-inflammatory; immunomodulator;  
 XX anti-infective; vaccine; anti-inflammatory; immunomodulator;  
 XX neuroprotective; anti-infective; vaccine; anti-inflammatory; immunomodulator;  
 XX expressed sequence tag; gene; ss.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200222660-A2.  
 XX  
 XX 21-MAR-2002.  
 XX  
 XX 10-SEP-2001; 2001WO-US026015.  
 XX  
 XX 11-SEP-2000; 2000US-00659671.  
 XX

XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT,  
 XX WPI: 2002-292408/33.  
 DR P-PSDB; ABB97283.  
 XX An isolated polynucleotide for treating diseases associated with its  
 PT encoded polypeptide such as cancer and multiple sclerosis.  
 PS Claim 1; SEQ ID NO 107; 509pp; English.  
 XX The present invention provides the protein and coding sequences of 444  
 CC novel human proteins. These were isolated from expressed sequences tags  
 CC (ESTs). They can be used to stimulate cell growth, to regulate  
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke  
 CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.  
 CC rheumatoid arthritis, and to treat nervous system disorders e.g.  
 CC Parkinson's disease. The present sequence is a coding sequence of the  
 CC invention  
 CC  
 SQ Sequence 5042 BP; 1448 A; 1034 C; 1199 G; 1361 T; 0 U; 0 Other;  
 Query Match 86.7%; Score 15.6; DB 6; Length 5042;  
 Best Local Similarity 83.3%; Pred. No. 7.3e+02;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GACCACTGTCGCGCATG 18  
 |||||:|||||  
 Db 149 GACCACTGTCGCGCATG 166  
 RESULT 20  
 ADE06975 ADE06975 standard; DNA; 5362 BP.  
 XX ADE06975;  
 AC ADE06975;  
 XX 29-JAN-2004 (first entry)  
 DT Novel coding sequence (useful for identifying genetic disorders) #41.  
 DE novel gene; novel protein; tissue marker; molecular weight marker;  
 XX chromosome marker; genetic disorder; gene; ds.  
 KM Unidentified.  
 XX  
 OS WO2003054152-A2.  
 PN  
 XX  
 PD 03-JUL-2003.  
 XX  
 PF 10-DEC-2002; 2002WO-US039555.  
 XX  
 PR 10-DEC-2001; 2001US-0339739P.  
 XX  
 PR 11-DEC-2001; 2001US-0339453P.  
 XX  
 PR 14-MAR-2002; 2002US-0365091P.  
 XX  
 PR 14-MAR-2002; 2002US-0365384P.  
 XX  
 PR 12-APR-2002; 2002US-0372381P.  
 XX  
 PR 12-APR-2002; 2002US-0372615P.  
 XX  
 PR 22-APR-2002; 2002US-0012855P.  
 XX  
 PR 24-APR-2002; 2002US-0376045P.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PT Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
 PI Ghosh W, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;  
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;  
 XX

DR WPI: 2003-569235/53.  
 DR P-PSDB; ADE07886.  
 XX New polynucleotides, useful for expressing recombinant proteins for  
 PT analysis, characterization or therapeutic use, or as markers for tissues  
 PT in which the corresponding protein is preferentially expressed.  
 XX  
 PS Claim 1; SEQ ID NO 41; 1177pp; English.  
 XX The invention comprises the amino acid and coding sequences of novel  
 CC proteins. The DNA and protein sequences of the invention are useful as:  
 CC markers for tissues in which the corresponding protein is preferentially  
 CC expressed; as molecular weight markers on gels; as chromosome markers or  
 CC tags; to identify chromosomes or to map related gene positions; and to  
 CC compare with endogenous DNA sequences in patients to identify potential  
 CC genetic disorders. The present DNA sequence represents a gene of the  
 CC invention.  
 CC  
 SQ Sequence 5362 BP; 1522 A; 1122 C; 1278 G; 1440 T; 0 U; 0 Other;  
 Query Match 86.7%; Score 15.6; DB 10; Length 5362;  
 Best Local Similarity 83.3%; Pred. No. 7.3e+02;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GACCACTGTCGCGCATG 18  
 |||||:|||||  
 Db 70 GACCACTGTCGCGCATG 87  
 RESULT 21  
 AAH89920 AAH89920 standard; cDNA; 5624 BP.  
 XX AAH89920;  
 AC AAH89920;  
 XX 01-OCT-2001 (first entry)  
 DT Human bone marrow cDNA, SEQ ID NO: 51.  
 XX  
 DE Human; bone marrow; antiinflammatory; cyostatic; neuroprotective;  
 XX antiviral; antibacterial; antifungal; anti-HIV; haemostatic;  
 XX immunosuppressive; gene therapy; cytokine cell proliferation;  
 KM cell differentiation modulator; immune disorder; infection; cancer;  
 KM human immunodeficiency virus; HIV; autoimmune disorder; haemophilia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153453-A2.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 23-DEC-2000; 2000WO-US034960.  
 XX  
 PR 23-DEC-1999; 99US-00471275.  
 XX  
 PR 21-JAN-2000; 2000US-00488725.  
 XX  
 PR 25-APR-2000; 2000US-0052317.  
 XX  
 PR 20-JUN-2000; 2000US-00598042.  
 XX  
 PR 19-JUL-2000; 2000US-00620312.  
 XX  
 PR 03-AUG-2000; 2000US-00653450.  
 XX  
 PR 14-SEP-2000; 2000US-00662191.  
 XX  
 PR 19-OCT-2000; 2000US-00693036.  
 XX  
 PR 30-NOV-2000; 2000US-0250583P.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PT Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;  
 PI Ren F, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Drmanac RT;  
 XX  
 DR WPI: 2001-488707/53.  
 DR P-PSDB; AAM00801.  
 XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful for

PT treating e.g. cancer and immune deficiency disorders.  
XX  
PS Claim 1, Page 238-240; 649pp; English.  
XX  
CC The present sequence is one of 251 novel human polynucleotides expressed in the bone marrow. The polynucleotide and the polypeptide encoded by it are useful in the treatment of various immune deficiencies and disorders. The deficiencies and disorders may be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal infection, or may result from an autoimmune disorder, a coagulation disorder (e.g. haemophilia), inhibition of tumour cell proliferation, suppression of an inflammatory response or treatment of a nervous system disorder such as Alzheimer's disease. Detection of the presence or increased expression of the polynucleotide or the protein it encodes is useful for the diagnosis and/or prognosis of one or more types of cancer. The polynucleotide and polypeptide can be used as nutritional sources or supplements and in the screening of chemical compounds as potential drugs  
XX  
SQ Sequence 5624 BP; 1555 A; 1205 C; 1349 G; 1515 T; 0 U; 0 Other;  
XX  
Query Match 86.7%; Score 15.6; DB 5; Length 5624;  
Best Local Similarity 83.3%; Pred. No. 7.2e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 GACCACTGTCGCGCATG 18  
85 GACCACTGTCGCGCATG 102  
Db  
RESULT 22  
ABLI0880/c  
ID ABLI0880 standard; cDNA; 12332 BP.  
XX  
AC ABLI0880;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27122.  
XX  
KM Drosophila; developmental biology; cell signalling; insecticide;  
XX  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
XX  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
DR P-PSDB; ABB66777.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.  
XX  
PS Claim 1; SEQ ID NO 27122; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABLI6176-ABLI0511), expressed DNA sequences (ABLI01840-ABLI6175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 12332 BP; 3452 A; 2580 C; 2685 G; 3615 T; 0 U; 0 Other;  
XX  
Query Match 86.7%; Score 15.6; DB 4; Length 12332;  
Best Local Similarity 83.3%; Pred. No. 7e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 GACCACTGTCGCGCATG 18  
1291 GTCCACTGTCGCGCATG 1274  
Db  
RESULT 23  
ABSS6454 01/c  
Continuation (2 of 22) of ABSS6454 from base 100001 (Streptococcus pneumoniae type 4 str.  
WP Sequence split into 22 fragments LOCUS ABSS6454 Accession ABSS6454  
WP Fragment Name Begin End  
WP ABSS6454\_00 1 110000  
WP ABSS6454\_01 100001 210000  
WP ABSS6454\_02 200001 310000  
WP ABSS6454\_03 300001 410000  
WP ABSS6454\_04 400001 510000  
WP ABSS6454\_05 500001 610000  
WP ABSS6454\_06 600001 710000  
WP ABSS6454\_07 700001 810000  
WP ABSS6454\_08 800001 910000  
WP ABSS6454\_09 900001 1010000  
WP ABSS6454\_10 1000001 1110000  
WP ABSS6454\_11 1100001 1210000  
WP ABSS6454\_12 1200001 1310000  
WP ABSS6454\_13 1300001 1410000  
WP ABSS6454\_14 1400001 1510000  
WP ABSS6454\_15 1500001 1610000  
WP ABSS6454\_16 1600001 1710000  
WP ABSS6454\_17 1700001 1810000  
WP ABSS6454\_18 1800001 1910000  
WP ABSS6454\_19 1900001 2010000  
WP ABSS6454\_20 2000001 2110000  
WP ABSS6454\_21 2100001 2162598  
Query Match 86.7%; Score 15.6; DB 10; Length 110000;  
Best Local Similarity 83.3%; Pred. No. 6.5e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 GACCACTGTCGCGCATG 18  
90126 GACCACTGTCGCGCATG 90109  
Db  
RESULT 24  
ABLI6420/c  
ID ABLI6420 standard; DNA; 5356 BP.  
XX  
AC ABLI6420;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 733.  
XX  
KM Drosophila; developmental biology; cell signalling; insecticide;  
XX  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX

PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 733; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 5356 BP; 1275 A; 1352 C; 1364 G; 1365 T; 0 U; 0 Other;  
SQ  
Query Match 84.4%; Score 15.2; DB 4; Length 5356;  
Best Local Similarity 87.5%; Pred. No. 1e+03;  
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 2 ACCAGTSCCGGCAT 17  
DB 377 ACCAGTSCCGGCAT 362  
RESULT 25  
ABD32859/C  
ID ABD32859 standard; DNA; 110021 BP.  
XX  
XX ABD32859;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Mouse cancer-associated genomic DNA MD17-042.  
XX  
XX Mouse; db; cancer-associated protein; gene; cytosolic; cancer;  
KM leukaemia; lymphoma; CAP.  
XX  
XX Mus musculus.  
OS  
XX WO2004074320-A2.  
PN  
XX 02-SEP-2004.  
PD  
XX 17-FEB-2004; 2004WO-US004730.  
PF  
XX 14-FEB-2003; 2003US-00367094.  
PR 14-MAR-2003; 2003US-00388838.  
PR 15-APR-2003; 2003US-00417375.  
PR 13-JUN-2003; 2003US-00461862.  
PR 15-SEP-2003; 2003US-00663431.  
PR 15-DEC-2003; 2003US-00737318.  
XX  
XX (SAGR-) SAGRES DISCOVERY INC.  
PA  
XX Morris DW, Morris DW, Malandro MS;  
PI  
XX WPI; 2004-652914/63.  
DR  
XX New isolated cancer-associated polynucleotides and polypeptides useful  
PT for diagnosing, preventing or treating cancers, especially lymphoma and  
PT

PT leukemia, or in screening for agents that modulate cancer.  
XX  
XX disclosure; seqid 533; 310pp; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid comprising at least 10  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
CC in the specification, or its complement. The nucleic acids encode cancer-  
CC associated proteins. Also included are an expression vector comprising  
CC the isolated nucleic acid cited above, a host cell comprising the above  
CC recombinant nucleic acid or expression vector, a microarray for detecting  
CC a cancer-associated (CA) nucleic acid comprising at least one probe  
CC comprising at least 10 contiguous nucleotides of any of the above-  
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
CC an open reading frame of a CA sequence selected from any of the 95  
CC polynucleotide sequences as mentioned in the specification, or its  
CC complement), an isolated antibody, (or its antigen binding fragment) that  
CC binds to the above polypeptide, a hybridoma that produces the above  
CC monoclonal antibody, a pharmaceutical composition comprising the above  
CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
CC cells (comprising the antibody cited above, methods for diagnosing cancer  
CC or for detecting the presence or absence of cancer cells in an  
CC individual, a method for inhibiting growth of cancer cells in an  
CC individual, a method for delivering a therapeutic agent to cancer cells  
CC in an individual, an electronic library comprising the above  
CC polynucleotide or polypeptide (or their fragments), methods of screening  
CC for anticancer activity or for a bioactive agent capable of modulating  
CC the activity of a CA protein (CAP), methods for detecting cancer  
CC associated with expression of a polypeptide in a test cell sample, a  
CC method for treating cancers and a method for inhibiting the expression of  
CC CA genes in a cell. The composition and methods are useful for detecting,  
CC diagnosing, preventing and treating cancer, especially lymphoma and  
CC leukaemia. These may also be used in screening for agents that modulate  
CC cancer. The present sequence is a mouse CAP genomic sequence. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 110021 BP; 27669 A; 24618 C; 25713 G; 30636 T; 0 U; 1385 Other;  
SQ  
Query Match 84.4%; Score 15.2; DB 13; Length 110021;  
Best Local Similarity 87.5%; Pred. No. 1e+03;  
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 3 CCAGTSCCGGCATG 18  
DB 41825 CCAGTSCCGGCATG 41810  
RESULT 26  
ABL7500  
ID ABL7500 standard; cDNA; 209 BP.  
XX  
XX ABL7500;  
XX  
DT 14-MAY-2002 (first entry)  
XX  
DE Corn tassell-derived polynucleotide (cdps) SEQ ID NO:4374.  
XX  
XX Corn; corn tassell-derived polynucleotide; cdps; hybrid breeding; CDPs;  
KM inheritance; characteristic; growth; development; disease resistance;  
KM environmental adaptability; quality; yield; molecular marker;  
KM multigene trait; plant breeding; corn tassell; gene; ss.  
XX  
XX Zea mays.  
OS  
XX US2001051335-A1.  
PN  
XX 13-DEC-2001.  
PD  
XX 16-APR-1999; 99US-00294093.  
PF  
XX 21-APR-1998; 98US-0082567P.  
PR  
XX

PA (LALG/) LALGUDI R V.  
 PA (ITOL/) ITO L Y.  
 PA (SHER/) SHERMAN B K.  
 XX Lalgudi RV, Ito LY, Sherman BK;  
 DR WPI; 2002-163647/21.  
 XX  
 PT Novel purified corn tassel-derived polynucleotide useful for determining  
 PT altered gene expression, to recover regulatory elements and to follow  
 PT inheritance of desirable characteristics through hybrid breeding  
 PT programs.  
 PS  
 PS Claim 1; SEQ ID NO 4374; 201pp; English.  
 XX  
 CC The present sequence describes a purified corn tassel-derived  
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence  
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences  
 CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (1)  
 CC can be used for determining altered gene expression, to recover  
 CC regulatory elements and to follow inheritance of desirable  
 CC characteristics through hybrid breeding programs. (1) are also useful in  
 CC the evaluation and alteration of desired characteristics associated with  
 CC growth and development, disease resistance, environmental adaptability,  
 CC quality and yield, and as molecular markers for studying inheritance of  
 CC multigene traits in a plant breeding program. (1) can be used to produce  
 CC a tassel-specific profile of gene transcription, a transcript image, to  
 CC clone regulatory elements for use in transformation vectors, to express a  
 CC polypeptide, to identify, isolate or extend identical or related corn  
 CC tassel nucleic acid sequences from DNA libraries, in nucleic acid  
 CC hybridisation or amplification technologies, as query sequences to  
 CC determine homology of known sequences, as probe for use in Southern or  
 CC Northern hybridisation, and to identify the presence of and/or to  
 CC determine the degree of similarity between two (or more) nucleic acid  
 CC sequences  
 XX  
 SQ Sequence 209 BP; 52 A; 62 C; 49 G; 42 T; 0 U; 4 Other;  
 Query Match 81.1%; Score 14.6; DB 6; Length 209;  
 Best Local Similarity 82.4%; Pred. No. 2.4e+03;  
 Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GACCACTGTCGGGCGAT 17  
 Db 87 GACTACGTGCCCCGGCAT 103  
 RESULT 27  
 ADQ52850  
 ID ADQ52850 standard; DNA; 520 BP.  
 XX  
 AC ADQ52850;  
 XX  
 DT 21-OCT-2004 (first entry)  
 XX  
 DE Novel canine microarray-related DNA sequence SeqID4152.  
 XX  
 KW canine microarray; drug screening; toxicity assay;  
 KW environmental pollutant; cellular response; gene expression profile;  
 KW toxic response; liver necrosis; fatty liver disease;  
 KW protein adduct formation; hepatitis; dog; ds.  
 XX  
 OS Canis familiaris.  
 XX  
 PN WO2004063324-A2.  
 XX  
 PD 29-JUN-2004.  
 XX  
 PF 05-MAY-2003; 2003WO-US013853.  
 XX  
 PR 03-MAY-2002; 2002US-0377240P.  
 XX  
 PA (GENE-) GENE LOGIC INC.

PA (PRIZ ) PFIZER PROD INC.  
 XX  
 XX Diggs JC, Porter M, Wei T;  
 XX  
 DR WPI; 2004-561890/54.  
 XX  
 PT New isolated nucleic acid molecule, useful for drug screening and  
 PT toxicity assays or for assessing the impact, including toxicity, of a  
 PT compound, pharmaceutical agent or environmental pollutant on a cell or  
 PT living organism.  
 PS  
 PS Claim 1; SEQ ID NO 4152; 41pp; English.  
 XX  
 CC This invention is related to a novel isolated canine nucleic acid  
 CC sequences and the construction of canine microarrays containing a  
 CC significant portion of the canine genome. The isolated canine nucleic  
 CC acid sequences of the invention may be useful for drug screening and  
 CC toxicity assays. The invention is therefore useful for assessing the  
 CC impact, including toxicity, of a compound, pharmaceutical agent or  
 CC environmental pollutant on a cell or living organism. The methods are  
 CC useful for detecting genes that are up- or down-regulated in canines in a  
 CC disease state. The sequences are useful as diagnostic agents or markers  
 CC to detect a cellular response in a sample individually or as part of a  
 CC gene expression profile. It is also useful as a target for agents that  
 CC modulate gene expression or activity. The database is useful for  
 CC producing electronic Northern blots that allow the user to determine the cell  
 CC type or tissue in which a given gene is expressed and to allow  
 CC determination of the abundance or expression level of a given gene in a  
 CC particular tissue or cell. The methods are useful for determining the  
 CC similarity of a toxic response to one or more individual compounds. The  
 CC methods are useful for predicting at least one toxic response or the  
 CC likelihood that a compound or test agent will induce various specific  
 CC pathologies such as those of the liver (liver necrosis), fatty liver  
 CC disease, protein adduct formation or hepatitis), those of the kidney,  
 CC heart, brain or testes, or other pathologies associated with at least one  
 CC of the toxins. The methods are also useful for predicting or elucidating  
 CC the potential cellular pathways influenced, induced or modulated by the  
 CC compound or test agent due to the similarity of the expression profile  
 CC compared to the profile induced by a known toxin. The present sequence is  
 CC that of a canine DNA sequence which was claimed for use during the  
 CC production of a canine microarray of the invention.  
 XX  
 SQ Sequence 520 BP; 72 A; 197 C; 156 G; 76 T; 0 U; 19 Other;  
 Query Match 81.1%; Score 14.6; DB 13; Length 520;  
 Best Local Similarity 82.4%; Pred. No. 2.4e+03;  
 Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GACCACTGTCGGGCGAT 17  
 Db 162 GGCCACGTCCCGGGCAT 178  
 RESULT 28  
 AAC56488/c  
 ID AAC56488 standard; DNA; 521 BP.  
 XX  
 AC AAC56488;  
 XX  
 DT 25-JAN-2001 (first entry)  
 XX  
 DE Eucalyptus grandis transcription factor DNA sequence #359.  
 XX  
 KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
 KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;  
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.  
 XX  
 OS Eucalyptus grandis.  
 XX  
 PN WO200053724-A2.

BD 14-SEP-2000.  
XX  
XX 09-MAR-2000; 2000WO-US006112.  
XX  
XX 11-MAR-1999; 99US-00266513.  
XX 18-AUG-1999; 99US-0149485P.  
XX  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
PA (FLEET-) FLETCHER CHALLENGE FORESTS LTD.  
XX  
XX Wood M, McGrath A, Shenk MA, Glenn M;  
XX WPI; 2000-579369/54.  
XX  
XX New isolated polynucleotide encoding a plant transcription factor for  
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
XX having modified gene expression or modified activity of a polypeptide.  
XX  
XX Claim 1; Page 447; 747pp; English.  
XX  
XX The present invention relates to novel plant transcription factors from  
XX Eucalyptus grandis or Pinus radiata. The present sequence is the coding  
XX sequence for one such transcription factor. The transcription factor may  
XX be used to produce a plant having modified gene expression such as a  
XX woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
XX mahogany species or to modify the activity of a polypeptide in a plant.  
XX The transcription factors of the present invention are members from the  
XX following families of regulatory proteins: bZIP, bZIP family of G-box  
XX binding factors, basic helix-loop-helix zipper,  
XX homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2  
XX and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and  
XX MYB  
SQ Sequence 521 BP; 99 A; 197 C; 141 G; 84 T; 0 U; 0 Other;  
Query Match 81.1%; Score 14.6; DB 3; Length 521;  
Best Local Similarity 82.4%; Pred. No. 2.4e+03;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GACCAAGTSCSGGCAT 17  
DB 497 GACCAAGTCCCGGCGT 481  
RESULT 29  
AAC56107/c  
ID AAC56107 standard; DNA; 521 BP.  
XX  
XX AAC56107;  
XX  
XX 25-JAN-2001 (first entry)  
XX  
XX Eucalyptus grandis transcription factor DNA sequence #238.  
XX  
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
XX poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
XX homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;  
XX type 2 Cys2His2; CCAAT box element; MYB; ss.  
XX  
XX Eucalyptus grandis.  
XX  
XX WO200053724-A2.  
XX  
XX 14-SEP-2000.  
XX  
XX 09-MAR-2000; 2000WO-US006112.  
XX  
XX 11-MAR-1999; 99US-00266513.  
XX 18-AUG-1999; 99US-0149485P.  
XX  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
XX (FLEET-) FLETCHER CHALLENGE FORESTS LTD.  
PA

XX Wood M, McGrath A, Shenk MA, Glenn M;  
XX WPI; 2000-579369/54.  
XX  
XX New isolated polynucleotide encoding a plant transcription factor for  
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
XX having modified gene expression or modified activity of a polypeptide.  
XX  
XX Claim 1; Page 109; 747pp; English.  
XX  
XX The present invention relates to novel plant transcription factors from  
XX Eucalyptus grandis or Pinus radiata. The present sequence is the coding  
XX sequence for one such transcription factor. The transcription factor may  
XX be used to produce a plant having modified gene expression such as a  
XX woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
XX mahogany species or to modify the activity of a polypeptide in a plant.  
XX The transcription factors of the present invention are members from the  
XX following families of regulatory proteins: bZIP, bZIP family of G-box  
XX binding factors, basic helix-loop-helix zipper,  
XX homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2  
XX and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and  
XX MYB  
SQ Sequence 521 BP; 99 A; 197 C; 141 G; 84 T; 0 U; 0 Other;  
Query Match 81.1%; Score 14.6; DB 3; Length 521;  
Best Local Similarity 82.4%; Pred. No. 2.4e+03;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GACCAAGTSCSGGCAT 17  
DB 497 GACCAAGTCCCGGCGT 481  
RESULT 30  
ADA48826  
ID ADA48826 standard; DNA; 568 BP.  
XX  
XX ADA48826;  
XX  
XX 20-NOV-2003 (first entry)  
XX  
XX Banana gene conferring disease resistance in plants.  
XX disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant;  
XX banana.  
XX  
XX Musa sp.  
XX  
XX WO200300906-A2.  
XX  
XX 03-JAN-2003.  
XX  
XX 21-JUN-2002; 2002WO-1B002453.  
XX  
XX 22-JUN-2001; 2001US-0300112P.  
XX 26-SEP-2001; 2001US-0352277P.  
XX 22-MAR-2002; 2002US-0366535P.  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;  
XX Katagiri F, Kreps J, Provart N, Rieke D, Zhu T;  
XX WPI; 2003-184052/18.  
XX  
XX New polynucleotide comprising a plant nucleotide sequence having an open  
XX reading frame that encodes a polypeptide associated with disease  
XX resistance, useful for conferring resistance or tolerance to a plant  
XX pathogen.  
XX  
XX Disclosure; SEQ ID NO 896; 299pp; English.  
PS

```

XX  The invention relates to a novel isolated polynucleotide comprising a
CC  plant nucleotide sequence having an open reading frame that encodes a
CC  polypeptide associated with disease resistance or its fragment having
CC  substantially the same activity as the full-length polypeptide. The
CC  polynucleotide of the invention is useful for conferring resistance or
CC  tolerance to a plant pathogen. The present sequence represents a gene
CC  conferring disease resistance used in the invention.
XX
SQ  Sequence 568 BP; 145 A; 154 C; 136 G; 128 T; 0 U; 5 Other;
    Query March      81.1%; Score 14.6; DB 9; Length 568;
    Best Local Similarity 82.4%; Pred. No. 2.4e+03;
    Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY  1 GACCACGTCGCGCAT 17
    |||||:|||||
Db  378 GACCACGTCGCGCAT 394

Search completed: July 20, 2005, 16:11:31
Job time : 331.143 secs

```

**THIS PAGE IS BLANK**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 14:42:10 ; Search time 852.429 Seconds  
(without alignments)  
1023.187 Million cell updates/sec

Title: US-10-017-471b-1  
Perfect score: 18  
Sequence: 1 gaccacgtaccggcagc 18

Scoring table: IDENTITY NUC  
Gap 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 300 summaries

Database : GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_scs.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.2	95.6	1203	1	STMAFSA
2	17.2	95.6	12070	1	AB011413
3	17.2	95.6	155628	1	AE016876
4	17.2	95.6	296300	1	AE005035
5	16	88.9	1215	1	AB001608
6	16	88.9	2234	1	AB001683
7	16	88.9	6727	1	SC0007731
8	16	88.9	290850	1	SC039127
9	15.6	86.7	513	6	BD139044
10	15.6	86.7	513	6	ARI98867
11	15.6	86.7	744	6	CQ585804
12	15.6	86.7	762	6	BD139042
13	15.6	86.7	762	6	ARI98865
14	15.6	86.7	859	3	AY060758
15	15.6	86.7	1008	6	AR480461
16	15.6	86.7	1188	6	AX567274
17	15.6	86.7	1408	6	E30250
18	15.6	86.7	1408	6	E30254
19	15.6	86.7	1408	6	AR215256

20	15.6	86.7	1408	6	AR215257	AR215257 Sequence
21	15.6	86.7	1798	8	NC425011	AR250111 Nitella c
22	15.6	86.7	2000	6	AX656460	AX656460 Sequence
23	15.6	86.7	2048	6	BD139028	BD139028 Virulence
24	15.6	86.7	2048	6	ARI98851	ARI98851 Sequence
25	15.6	86.7	2395	6	C0789114	C0789114 Sequence
26	15.6	86.7	2395	6	AR218973	AR218973 Sequence
27	15.6	86.7	2395	6	BD003885	BD003885 Polynucle
28	15.6	86.7	3233	9	HS112B	X81346 Homo sapien
29	15.6	86.7	4927	9	AF227905	AF227905 Homo sapi
30	15.6	86.7	5042	6	AX405692	AX405692 Sequence
31	15.6	86.7	6252	1	SVU24659	U24659 Streptomyce
32	15.6	86.7	6684	9	BC041098	BC041098 Homo sapi
33	15.6	86.7	10127	1	AE011135	AE011135 Methanosa
34	15.6	86.7	10643	1	AE005070	AE005070 Halobacte
35	15.6	86.7	11333	1	AE008410	AE008410 Streptoco
36	15.6	86.7	11790	1	AE011835	AE011835 Xanthomon
37	15.6	86.7	12332	6	C0585803	C0585803 Sequence
38	15.6	86.7	12606	1	AE005995	AE005995 Caulobact
39	15.6	86.7	13121	1	AE004608	AE004608 Pseudomon
40	15.6	86.7	13385	1	AE007344	AE007344 Streptoco
41	15.6	86.7	45752	3	DMBN33B1	AL133506 Drosophi
42	15.6	86.7	72284	2	AC020328	AC020328 Drosophi
43	15.6	86.7	77428	3	AC017634	AC017634 Drosophi
44	15.6	86.7	80389	3	AC005643	AC005643 Drosophi
45	15.6	86.7	103124	9	AL139818	AL139818 Human DNA
46	15.6	86.7	103714	2	AC146964	AC146964 Caroli
47	15.6	86.7	110000	1	AP006618	Continuation (28 o
48	15.6	86.7	112911	9	AC005865	AC005865 Homo sapi
49	15.6	86.7	115932	9	AC011446	AC011446 Homo sapi
50	15.6	86.7	124950	9	AC108059	AC108059 Homo sapi
51	15.6	86.7	131246	8	OSJND0299	EX42605 Oryza sat
52	15.6	86.7	150089	8	AC144743	AC144743 Oryza sat
53	15.6	86.7	154681	3	AC039908	AC039908 Drosophi
54	15.6	86.7	156573	2	AC073075	AC073075 Homo sapi
55	15.6	86.7	151569	3	AC104145	AC104145 Drosophi
56	15.6	86.7	163424	2	AC129293	AC129293 Mus muscu
57	15.6	86.7	165696	9	AL353803	AL353803 Human DNA
58	15.6	86.7	168519	9	AC144737	AC144737 Oryza sat
59	15.6	86.7	170929	10	AC130548	AC130548 Mus muscu
60	15.6	86.7	171222	2	AC091387	AC091387 Drosophi
61	15.6	86.7	175936	2	SPNBT1908	AL44930 Streptoco
62	15.6	86.7	176383	8	OSJND0003	AL606460 Oryza sat
63	15.6	86.7	178062	9	CNS01DRK	AL117356 Human chr
64	15.6	86.7	186607	2	AC140777	AC140777 Mus muscu
65	15.6	86.7	190177	2	AC009959	AC009959 Homo sapi
66	15.6	86.7	210614	1	AB088224	AB088224 Streptomy
67	15.6	86.7	210973	2	AC130949	AC130949 Rattus no
68	15.6	86.7	211397	2	AC151577	AC151577 Mus muscu
69	15.6	86.7	211829	10	AL645854	AL645854 Rattus no
70	15.6	86.7	215050	1	AL646057	AL646057 Rattus no
71	15.6	86.7	229788	2	AC095653	AC095653 Rattus no
72	15.6	86.7	229635	2	AC094256	AC094256 Rattus no
73	15.6	86.7	232042	2	AC128144	AC128144 Rattus no
74	15.6	86.7	240904	2	AC126204	AC126204 Rattus no
75	15.6	86.7	242135	2	AC127842	AC127842 Rattus no
76	15.6	86.7	247254	1	AC114844	AC114844 Rattus no
77	15.6	86.7	260050	1	SME591782	AL591782 Streptomy
78	15.6	86.7	265472	3	AE003817	AE003817 Drosophi
79	15.6	86.7	269267	1	AC110861	AC110861 Rattus no
80	15.6	86.7	299050	2	SC039104	AL939104 Streptomy
81	15.6	86.7	309052	3	AE003425	AE003425 Drosophi
82	15.6	86.7	347786	1	EX640438	EX640438 Bordetell
83	15.6	86.7	349146	1	EX640424	EX640424 Bordetell
84	15.6	86.7	349980	6	AX571760	AX571760 Sequence
85	15.6	84.4	3284	1	AB117719	AB117719 Rhodococc
86	15.6	84.4	5356	6	C0594113	C0594113 Sequence
87	15.2	84.4	94874	10	AL808126	AL808126 Mouse DNA
88	15.2	84.4	109151	6	AC017637	AC017637 Drosophi
89	15.2	84.4	110021	6	C0870112	C0870112 Sequence
90	15.2	84.4	161278	3	AC008223	AC008223 Drosophi
91	15.2	84.4	181021	10	AL663031	AL663031 Mouse DNA
92	15.2	84.4	184889	9	AL442125	AL442125 Human DNA

C 93	15.2	84.4	185639	2	AC145465	AC145465 Callipectu	C 166	14.6	81.1	2567	10	AK098097	AK098097 Mus muscu
C 94	15.2	84.4	190155	3	AC008201	AC008201 Drosophila	C 167	14.6	81.1	2761	9	BC043916	BC043916 Homo sapi
C 95	15.2	84.4	208880	2	AC128103	AC128103 Rattus no	C 168	14.6	81.1	2879	6	C0722672	C0722672 Sequence
C 96	15.2	84.4	239972	2	AC095835	AC095835 Rattus no	C 169	14.6	81.1	2883	1	SY0ATBP	D14438 Synechococ
C 97	15.2	84.4	244416	2	AE003744	AE003744 Drosophila	C 170	14.6	81.1	2996	6	C0580250	C0580250 Sequence
C 98	15.2	84.4	250902	2	AC131965	AC131965 Rattus no	C 171	14.6	81.1	3290	3	DME011320	AE011320 Drosophila
C 99	15.2	84.4	292200	1	SC0939129	SC0939129 Streptomy	C 172	14.6	81.1	3375	8	AF169793	AF169793 Podospira
C 100	15.2	84.4	299925	1	AP005045	AP005045 Streptomy	C 173	14.6	81.1	3380	6	C0580244	C0580244 Sequence
C 101	15.2	84.4	300140	1	AP005556	AP005556 Bradyrhiz	C 174	14.6	81.1	3500	6	C0572837	C0572837 Sequence
C 102	15.2	84.4	300297	1	AE016944	AE016944 Bacteroid	C 175	14.6	81.1	3542	1	AF039105	AF039105 Sequence
C 103	15	83.3	309	13	AF427374	AF427374 Unculture	C 176	14.6	81.1	3630	9	HSMB04113	HSMB04113 Homo sapi
C 104	15	83.3	309	13	AF427376	AF427376 Unculture	C 177	14.6	81.1	4054	3	MSU50719	MSU50719 Manduca sex
C 105	15	83.3	309	13	AF427378	AF427378 Unculture	C 178	14.6	81.1	4360	6	C0602123	C0602123 Sequence
C 106	15	83.3	576	11	BY030613	BY030613 S212P6013	C 179	14.6	81.1	4421	9	HUMCATP	L06620 Human plasm
C 107	15	83.3	39737	9	AC011521	AC011521 Homo sapi	C 180	14.6	81.1	4789	9	AE046776	AE046776 Homo sapi
C 108	15	83.3	48109	2	AC151741	AC151741 Rattus no	C 181	14.6	81.1	4979	9	HUMATP282X	L09777 Human plasm
C 109	15	83.3	66256	2	AC087454	AC087454 Homo sapi	C 182	14.6	81.1	5137	14	AY044867	AY044867 Hepatitis
C 110	15	83.3	72229	2	AC101639	AC101639 Mus muscu	C 183	14.6	81.1	6037	10	AK122256	AK122256 Mus muscu
C 111	15	83.3	135975	9	AC105935	AC105935 Homo sapi	C 184	14.6	81.1	6676	1	AF031590	AF031590 Streptomy
C 112	15	83.3	135439	9	AC008626	AC008626 Homo sapi	C 185	14.6	81.1	6755	1	AF002247	AF002247 Rhodococc
C 113	15	83.3	163424	2	AC099330	AC099330 Homo sapi	C 186	14.6	81.1	6946	6	AX818197	AX818197 Sequence
C 114	15	83.3	167471	9	AC019211	AC019211 Homo sapi	C 187	14.6	81.1	6946	9	AB007919	AB007919 Homo sapi
C 115	15	83.3	167722	9	AC073548	AC073548 Homo sapi	C 188	14.6	81.1	7838	6	C0601595	C0601595 Sequence
C 116	15	83.3	174894	9	AC022145	AC022145 Homo sapi	C 189	14.6	81.1	8246	1	AE014770	AE014770 Bifidobac
C 117	15	83.3	187723	9	AC016889	AC016889 Homo sapi	C 190	14.6	81.1	8324	2	AC019926	AC019926 Drosophila
C 118	15	83.3	189149	2	AC046171	AC046171 Homo sapi	C 191	14.6	81.1	8931	1	AF161183	AF161183 Burkholde
C 119	15	83.3	204746	2	AC055866	AC055866 Homo sapi	C 192	14.6	81.1	9198	1	AB002355	AB002355 Homo sapi
C 120	15	83.3	217325	10	AC115360	AC115360 Mus muscu	C 193	14.6	81.1	10023	1	AE009262	AE009262 Agrobacte
C 121	15	83.3	239718	2	AC101774	AC101774 Mus muscu	C 194	14.6	81.1	10235	1	AE008350	AE008350 Agrobacte
C 122	15	83.3	252486	2	AC111445	AC111445 Rattus no	C 195	14.6	81.1	10769	1	AE000881	AE000881 Methanoba
C 123	15	83.3	253946	2	AC105161	AC105161 Mus muscu	C 196	14.6	81.1	10970	1	AE001869	AE001869 Deinococc
C 124	15	83.3	254939	2	AC099076	AC099076 Rattus no	C 197	14.6	81.1	11188	6	AR339644	AR339644 Sequence
C 125	15	83.3	273325	4	AF324420	AF324420 Equus cab	C 198	14.6	81.1	11188	6	AR275002	AR275002 Sequence
C 126	14.6	81.1	465	9	HS038572	HS038572 Homo sapi	C 199	14.6	81.1	12198	1	AF187159	AF187159 Streptomy
C 127	14.6	81.1	510	6	AR425418	AR425418 Sequence	C 200	14.6	81.1	12404	1	AE014736	AE014736 Bifidobac
C 128	14.6	81.1	510	6	AX986112	AX986112 Sequence	C 201	14.6	81.1	13221	1	AE014422	AE014422 Brucella
C 129	14.6	81.1	510	6	BD120971	BD120971 EST and e	C 202	14.6	81.1	13453	1	AE009518	AE009518 Brucella
C 130	14.6	81.1	568	6	AX660539	AX660539 Sequence	C 203	14.6	81.1	13537	9	AE257737	AE257737 Homo sapi
C 131	14.6	81.1	635	5	AY394439	AY394439 Spheroelid	C 204	14.6	81.1	13816	9	HS040468	HS040468 Homo sapi
C 132	14.6	81.1	766	8	AF051369	AF051369 Oryza sat	C 205	14.6	81.1	14069	2	AC019558	AC019558 Drosophila
C 133	14.6	81.1	808	11	G06627	G06627 human STS W	C 206	14.6	81.1	14704	1	AE001970	AE001970 Deinococc
C 134	14.6	81.1	808	11	G10673	AR123908 Sequence	C 207	14.6	81.1	15836	1	AE005057	AE005057 Halobacte
C 135	14.6	81.1	888	8	AR123908	AK119706 Oryza sat	C 208	14.6	81.1	17083	6	AX536588	AX536588 Streptomy
C 136	14.6	81.1	913	8	AK119706	AK119706 Oryza sat	C 209	14.6	81.1	17083	6	AX593213	AX593213 Sequence
C 137	14.6	81.1	915	8	AK064673	AK064673 Oryza sat	C 210	14.6	81.1	18524	6	CQ730656	CQ730656 Sequence
C 138	14.6	81.1	921	6	C0580251	C0580251 Sequence	C 211	14.6	81.1	20435	9	HS042535	HS042535 Homo sapi
C 139	14.6	81.1	922	6	AY051697	AY051697 Drosophila	C 212	14.6	81.1	21051	2	AC017802	AC017802 Drosophila
C 140	14.6	81.1	1004	10	AF056244	AF056244 Mus muscu	C 213	14.6	81.1	28222	2	AC019963	AC019963 Drosophila
C 141	14.6	81.1	1005	6	CQ445545	CQ445545 Sequence	C 214	14.6	81.1	34667	1	AY233211	AY233211 Streptomy
C 142	14.6	81.1	1041	6	AX573922	AX573922 Sequence	C 215	14.6	81.1	34667	1	HS0344905	HS0344905 Homo sapi
C 143	14.6	81.1	1068	6	AX55318	AX55318 Sequence	C 216	14.6	81.1	36615	1	AF449411	AF449411 Myxococcu
C 144	14.6	81.1	1173	10	NM2427265	NM2427265 Masourtie	C 217	14.6	81.1	37108	3	CET20F10	CET20F10 Caenothabdi
C 145	14.6	81.1	1215	3	U00790	U00790 Drosophila	C 218	14.6	81.1	45055	6	AX574197	AX574197 Sequence
C 146	14.6	81.1	1242	3	CQ572838	CQ572838 Sequence	C 219	14.6	81.1	49899	10	AC134735	AC134735 Rattus no
C 147	14.6	81.1	1242	3	AR510109	AR510109 Sequence	C 220	14.6	81.1	50000	6	AX535264	AX535264 Sequence
C 148	14.6	81.1	1435	8	AK105163	AK105163 Oryza sat	C 221	14.6	81.1	50000	6	AX535266	AX535266 Sequence
C 149	14.6	81.1	1435	8	BC072173	BC072173 Xenopus 1	C 222	14.6	81.1	59816	1	AF333038	AF333038 Streptomy
C 150	14.6	81.1	1630	5	BC012206	BC012206 Mus muscu	C 223	14.6	81.1	61125	2	AC137716	AC137716 Homo sapi
C 151	14.6	81.1	1649	10	PYU19903	PYU19903 Prairie vol	C 224	14.6	81.1	61371	2	AC105135	AC105135 Homo sapi
C 152	14.6	81.1	1722	14	AR339656	AR339656 Sequence	C 225	14.6	81.1	62001	2	AC015206	AC015206 Drosophila
C 153	14.6	81.1	1749	6	AX275014	AX275014 Sequence	C 226	14.6	81.1	82972	2	NCB24H17	NCB24H17 Homo sapi
C 154	14.6	81.1	1749	6	AX275014	AX275014 Sequence	C 227	14.6	81.1	82972	8	AC023162	AC023162 Homo sapi
C 155	14.6	81.1	1793	10	AF031035	AF031035 Mus muscu	C 228	14.6	81.1	95579	9	HS043402	HS043402 Homo sapi
C 156	14.6	81.1	1898	1	SCTRXABGN	SCTRXABGN X92105 Streptomyce	C 229	14.6	81.1	98356	9	AC105210	AC105210 Homo sapi
C 157	14.6	81.1	1931	6	BD179569	BD179569 Highly th	C 230	14.6	81.1	101777	9	AC008116	AC008116 Homo sapi
C 158	14.6	81.1	2394	6	BD179569	BD179569 Highly th	C 231	14.6	81.1	104947	9	AE000516	AE000516_34
C 159	14.6	81.1	2397	1	D86177	D86177 Mouse mRNA	C 232	14.6	81.1	110000	1	AE016822_24	AE016822_24
C 160	14.6	81.1	2519	6	C0841741	C0841741 Sequence	C 233	14.6	81.1	110000	1	AE017180_11	AE017180_11
C 161	14.6	81.1	2519	6	AK123359	AK123359 Homo sapi	C 234	14.6	81.1	110000	1	AE017282_22	AE017282_22
C 162	14.6	81.1	2534	9	AK834822	AK834822 Sequence	C 235	14.6	81.1	110000	1	AP006840_01	AP006840_01
C 163	14.6	81.1	2534	9	AK097489	AK097489 Homo sapi	C 236	14.6	81.1	110000	1	AP006840_02	AP006840_02
C 164	14.6	81.1	2534	10	BC031774	BC031774 Mus muscu	C 237	14.6	81.1	110000	1	AP006840_02	AP006840_02
C 165	14.6	81.1	2534	10	BC031774	BC031774 Mus muscu	C 238	14.6	81.1	110000	1	AP006840_26	AP006840_26

AK098097 Mus muscu  
BC043916 Homo sapi  
C0722672 Sequence  
D14438 Synechococ  
C0580250 Sequence  
AE011320 Drosophila  
AF169793 Podospira  
C0580244 Sequence  
C0572837 Sequence  
AF039105 Methanoba  
AL832802 Homo sapi  
US071971 Manduca sex  
C0602123 Sequence  
L06620 Human plasm  
AE046776 Homo sapi  
L09777 Human plasm  
AY044867 Hepatitis  
AK122256 Mus muscu  
AF031590 Streptomy  
AF002247 Rhodococc  
AX818197 Sequence  
AB007919 Homo sapi  
C0601595 Sequence  
AE014770 Bifidobac  
AC019926 Drosophila  
AF161183 Burkholde  
AB002355 Homo sapi  
AE009262 Agrobacte  
AE008350 Agrobacte  
AE000881 Methanoba  
AE001869 Deinococc  
AR339644 Sequence  
AR275002 Sequence  
AF187159 Streptomy  
AE014736 Bifidobac  
AE014422 Brucella  
AE009518 Brucella  
AE257737 Homo sapi  
AF404468 Homo sapi  
AC019558 Drosophila  
AE001970 Deinococc  
AE005057 Halobacte  
AX536588 Streptomy  
AX593213 Sequence  
CQ730656 Sequence  
AF002355 Homo sapi  
AC017802 Drosophila  
AC019963 Drosophila  
AY233211 Streptomy  
AF144805 Homo sapi  
AF449411 Myxococcu  
Z81594 Caenothabdi  
AX574197 Sequence  
AC134735 Rattus no  
AX535264 Sequence  
AX535266 Sequence  
AF333038 Streptomy  
AC137716 Homo sapi  
AC105135 Homo sapi  
AC015206 Drosophila  
AC013683 Homo sapi  
AL356615 Neurospor  
AC023162 Homo sapi  
AL449421 Homo sapi  
AC105210 Homo sapi  
AC008116 Homo sapi  
Continuation (35 o  
Continuation (25 o  
Continuation (12 o  
Continuation (23 o  
Continuation (2 of  
Continuation (3 of  
Continuation (27 o

C 239	14.6	81.1	110000	2	AC096315.6	Continuation (7 of
240	14.6	81.1	110000	2	AC15182_1	Continuation (2 of
241	14.6	81.1	110000	2	AC151851_0	AC151851 Takifugu
242	14.6	81.1	110000	2	AC151851_1	Continuation (2 of
243	14.6	81.1	110000	2	AP006494.0	AP006494 Cyanidios
244	14.6	81.1	110000	2	AP006500_00	AP006500 Cyanidios
245	14.6	81.1	110000	8	CR382121_05	Continuation (6 of
246	14.6	81.1	110000	8	CR382128_15	Continuation (16 of
247	14.6	81.1	115968	9	AL139100	AL139100 Homo sapi
248	14.6	81.1	118699	9	AC133004	AC133004 Homo sapi
249	14.6	81.1	121849	8	AC092779	AC092779 Oryza sat
250	14.6	81.1	126784	2	AY263454	AY263454 Sus scrofa
251	14.6	81.1	128000	9	AC098690	AC098690 Homo sapi
C 252	14.6	81.1	128638	9	AC015968	AC015968 Homo sapi
C 253	14.6	81.1	135964	9	AL353593	AL353593 Homo sapi
C 254	14.6	81.1	137952	2	AC105141	AC105141 Homo sapi
C 255	14.6	81.1	142329	2	AC128499	AC128499 Rattus no
C 256	14.6	81.1	146947	2	AC012515	AC012515 Homo sapi
C 257	14.6	81.1	148675	10	AC131769	AC131769 Mus muscu
C 258	14.6	81.1	153648	3	AC008210	AC008210 Drosophi
C 259	14.6	81.1	154110	2	AC148737	AC148737 Ornitho
C 260	14.6	81.1	154736	2	AL139246	AL139246 Homo sapi
C 261	14.6	81.1	158892	2	CNS01R14	AL162871 Homo sapi
C 262	14.6	81.1	159880	10	AC142258	AC142258 Mus muscu
C 263	14.6	81.1	161096	9	AC016257	AC016257 Homo sapi
C 264	14.6	81.1	161117	3	AC008311	AC008311 Drosophi
C 265	14.6	81.1	162829	3	CR352283	CR352283 Dario rer
C 266	14.6	81.1	163455	2	AC009427	AC009427 Homo sapi
C 267	14.6	81.1	163579	9	AC093563	AC093563 Homo sapi
C 268	14.6	81.1	164396	2	HSAC00380	AC0000380 Homo sapi
C 269	14.6	81.1	164766	2	AC026657	AC026657 Homo sapi
C 270	14.6	81.1	164959	2	AC087823	AC087823 Homo sapi
C 271	14.6	81.1	166007	2	CNS01DRC	AL117186 Human chr
C 272	14.6	81.1	166176	2	AC104143	AC104143 Drosophi
C 273	14.6	81.1	166176	2	AC104143	AC104143 Drosophi
C 274	14.6	81.1	166450	2	AC078874	AC078874 Homo sapi
C 275	14.6	81.1	167601	2	AC120208	AC120208 Canis fam
C 276	14.6	81.1	168448	3	AC009346	AC009346 Drosophi
C 277	14.6	81.1	169168	5	BX546502	BX546502 Zebrafish
C 278	14.6	81.1	169189	2	AC125293	AC125293 Drosophi
C 279	14.6	81.1	171032	2	AC091761	AC091761 Felis cat
C 280	14.6	81.1	171421	8	AP004776	AP004776 Oryza sat
C 281	14.6	81.1	172560	2	AC091794	AC091794 Felis cat
C 282	14.6	81.1	173441	9	AP000787	AP000787 Homo sapi
C 283	14.6	81.1	173702	3	AC007593	AC007593 Drosophi
C 284	14.6	81.1	174234	10	AC117098	AC117098 Rattus no
C 285	14.6	81.1	174612	2	AC023889	AC023889 Homo sapi
C 286	14.6	81.1	174673	9	AC019187	AC019187 Homo sapi
C 287	14.6	81.1	174906	9	AC067930	AC067930 Homo sapi
C 288	14.6	81.1	175996	2	AC108480	AC108480 Drosophi
C 289	14.6	81.1	176205	2	AC150984	AC150984 Bos tauru
C 290	14.6	81.1	176918	2	AC109613	AC109613 Bos tauru
C 291	14.6	81.1	180727	2	AC021174	AC021174 Homo sapi
C 292	14.6	81.1	180783	9	AC079611	AC079611 Homo sapi
C 293	14.6	81.1	181371	2	AC018893	AC018893 Homo sapi
C 294	14.6	81.1	181515	2	AC119068	AC119068 Canis fam
C 295	14.6	81.1	181904	3	AC008211	AC008211 Drosophi
C 296	14.6	81.1	182785	9	AC145937	AC145937 Pan trogl
C 297	14.6	81.1	186431	10	AC113092	AC113092 Mus muscu
C 298	14.6	81.1	188549	3	AC007818	AC007818 Drosophi
C 299	14.6	81.1	189050	1	AL646069	AL646069 Ralstonia
C 300	14.6	81.1	189317	9	HSJ190720	AL050335 Human DNA

## ALIGNMENTS

RESULT 1  
STMAFSA  
LOCUS  
DEFINITION  
ACCESSION

STMAFSA  
S.grieseus afesa gene encoding a possible A-factor biosynthesis  
protein.  
M24250

1203 bp DNA linear BCT 26-APR-1993

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source

M24250.1 GI:153148  
A-factor biosynthesis.  
Streptomyces grieseus  
Streptomyces grieseus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomyces; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 1203)  
Horinouchi, S., Suzuki, H., Nishiyama, M. and Bappu, T.  
Nucleotide sequence and transcriptional analysis of the  
Streptomyces grieseus gene (afsa) responsible for A-factor  
biosynthesis  
J. Bacteriol. 171 (2), 1206-1210 (1989)  
89123125  
2492509  
Original source text: S.grieseus DNA.  
Location/Qualifiers  
1..1203  
/organism="Streptomyces grieseus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1911"  
204..1109  
/note="afsa protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AAA26693.1"  
/db\_xref="GI:153149"  
/translation="MDAAEVAHPVGIEMVHTRPEDAFRNWRIGRDFAVEALP  
VDHFPFAPVGDLDHDLVLAEMARQAMLAFFAGYGIPIGYHFLITLIDVCHPEHIG  
VGGPEPIGEVFCSDLKMRAGIPAGQGVMAVGRGDRLAATGVAATRFSTPAVRM  
RGDPVPEGISLPETAVPAPSPAGRAGVAVEDVSTGREGVETLRVDRHTLRFORPD  
HYPMLEBARQACLVAGPAGVIEVETATRTFRHSERSSPCWIGAVVPGADEDTV  
TVATGHQDEIVFTVLSGPRAGH"

ORIGIN

Query Match 95.6%; Score 17.2; DB 1; Length 1203;  
Best Local Similarity 88.9%; Pred. No. 1.4e+03;  
Matches 16; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Cy 1 GACCACGTCSCGSGCATG 18  
Db 855 GACCACGTCGCGGCATG 872

RESULT 2  
AB011413  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source

AB011413 12070 bp DNA linear BCT 07-AUG-1998  
Streptomyces grieseus genes for Orf2, Orf3, Orf4, Orf5, AfesA, Orf8,  
partial and complete cds.  
AB011413  
AB011413.1 GI:3401946  
Orf8; AfesA; Orf5; Orf4; Orf3; Orf2.  
Streptomyces grieseus  
Streptomyces grieseus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomyces; Streptomycetaceae; Streptomyces.  
1 (sites)  
Uneyama, T.  
Open reading frame encoded around afesA gene  
Unpublished  
2 (bases 1 to 12070)  
Uneyama, T.  
Direct Submission  
Submitted (23-FEB-1998) Takashi Uneyama, University of Tokyo,  
Department of Agriculture and Life Sciences, Yayoi 1-1-1,  
Bunkyo-Ku, Tokyo 113, Japan  
(E-mail:aa67103@hongo.ecc.u-tokyo.ac.jp, Tel:+81-3-3812-2111)  
Location/Qualifiers  
1..12070  
/organism="Streptomyces grieseus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1911"  
complement(2660..3511)  
/gene="orf2"

Query Match	95.6%	Score 17.2	DB 1	Length 12070
Best Local Similarity	88.9%	Pred. No. 9e+02		
Matches 16	Conservative	2	Mismatches 0	Indels 0
			Gaps 0	
Db	8137	GACCACGTCGCCGACATG	8154	
Oy	1	GACCACGTCGCCGACATG	18	
		: : : :		
RESULT 3				
LOCUS	AE016876	155628 bp	DNA	linear BCT 09-OCT-2003
DEFINITION	Pseudomonas syringae pv. tomato str. DC3000 section 21 of 21 of the complete genome.			
ACCESSION	AE016876	AE016853		
VERSION	AE016876.1	GI:28855841		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS</				

[illegible]





receptor for an extracellular regulatory factor (IM-2) from  
Streptomyces sp. strain FRI-5  
JOURNAL J. Bacteriol. 179 (16), 5131-5137 (1997)  
MEDLINE 97405912  
PUBMED 9260956  
REFERENCE 2 (bases 1 to 2234)  
AUTHORS Waki,M.  
TITLE Direct Submission  
JOURNAL Submitted (07-MAR-1997) Miyoko Waki, Osaka University, Graduate  
School of Engineering, Department of Biotechnology; Yamadaoka 2-1,  
Suita, Osaka 565, Japan (E-mail:waki.yam@stu.bio.eng.osaka-u.ac.jp).  
Tel:+81-6-879-7433, Fax:+81-6-879-7432)  
location/Qualifiers  
FEATURES  
source  
1..2234  
/organism="Streptomyces sp."  
/mol\_type="genomic DNA"  
/strain="FRI-5"  
/db\_xref="taxon:1931"  
440..1315  
/gene="farX"  
440..1315  
/gene="farX"  
/codon\_start=1  
/transl\_except=(pos:440..442,aa:Met)  
/transl\_table=11  
/product="FarX"  
/protein\_id="BAA21858.1"  
/db\_xref="GI:2342429"  
/translation="MWHTSTAOVLITNDMRLDPAFSVTARWPLSHAFTPVGDGY  
DPLNCATRIROIATILGHAFRAVPGHGFVMDLSVYRPEILRVGLVPTVLAIT  
CVELIKRRAISGLSTEAIVRRDGVATGASATCTSPAYTORIPEHVITPPEHPL  
PLTAPAAQSVARLSPIDVLSPLDERNQRLVDTNHPVLFDHVDHVPQWVMEAA  
ROAASALGRPSFMPLGVAGSEFKRYVELDAPCVIESRLFDVDGAEVAVTGHONG  
ELTFVGYTASSYGY"  
1500..2165  
/gene="farA"  
1500..2165  
/gene="farA"  
/codon\_start=1  
/transl\_except=(pos:1500..1502,aa:Met)  
/transl\_table=11  
/product="FarA"  
/protein\_id="BAA21859.1"  
/db\_xref="GI:2342430"  
/translation="MAEQVRAIRTRQALISAARVDERGYQAATISILTVAGYTK  
ALYHPQKEDLAQGLTAONEDLLPERPAQLOEVDAVMHTHRLTNPVAGVR  
LSIDVNAAGLDRSAPRWWDKFTDLEKAAQGLPHVVPATADVITGAYGVOS  
MSQALTERHODIGQRVNALIRHIMPSIAQPSVLASLHGESRAEEVYLEARQIARQAD  
EED"  
ORIGIN  
Query Match 88.9%; Score 16; DB 1; Length 2234;  
Beet Local Similarity 88.9%; Pred.No. 4.6e+03;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
CY 1 GACCACTGSCCGGCATG 18  
|||||  
DB 1056 GACCACTGATCCGGCATG 1075  
RESULT 7  
SC0007731 6727 bp DNA linear BCT 24-AUG-1998  
LOCUS 07731/c  
DEFINITION Streptomyces coelicolor scbr gene, scbA gene, ORFs A,B,X & Z.  
ACCESSION AJ007731.1 GI:3425857  
VERSION  
KEYWORDS gamma-butyrolactone binding protein; scbA gene; scbr gene.  
SOURCE Streptomyces coelicolor A3(2)  
ORGANISM Streptomyces coelicolor A3(2)  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE  
1  
Takano,E., Chakraborty,R., Nihira,T., Yamada,Y. and Bibb,M.

TITLE  
JOURNAL Characterisation of scbr, and scbA genes involved in  
gamma-butyrolactone binding and synthesis in Streptomyces  
coelicolor  
REFERENCE 2 (bases 1 to 6727)  
AUTHORS Takano,E.  
TITLE Direct Submission  
JOURNAL Submitted (11-AUG-1998) Takano E., Dept. Genetics, John Innes  
Institute, Colney lane, Norwich, NR4 7UH, UK  
location/Qualifiers  
FEATURES  
source  
1..6727  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="M145"  
/db\_xref="taxon:100226"  
37..969  
/gene="orfB"  
37..969  
/gene="orfB"  
/codon\_start=1  
/transl\_table=1  
/product="histidine kinase"  
/protein\_id="CAA07625.1"  
/db\_xref="GI:3425858"  
/db\_xref="GOA:O86849"  
/db\_xref="UniProt/TREMBL:O86849"  
/translation="MTGPSPIENRRGLIPCSIDRPVSSRRTERHPCPRTPARD  
LAGSHRGGRFETLQDGLSRSPRARQITGTIAPGIPAPGRPPAGPVTDRGRD  
HVAMERRRLAREVDELTALSTPAHHIETLAAETGCTPLDAARSLRAEMVARRL  
TGELOAQTVPLPLAAVGEFPAATRPRTGTEARJRTTDEBRLSDVCRLEFLVAREAL  
HNARHNAAARVYTLARFPRMAAGIVDGVGPDAVALVPGHRAAGLSMMDRID  
VGMRLIVSGAGGTHIDVLPDRKYSTAPRT"  
complement(1000..1854)  
/gene="orfA"  
complement(1000..1854)  
/gene="orfA"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="CAA07626.1"  
/db\_xref="GI:3425859"  
/db\_xref="UniProt/TREMBL:O86850"  
/translation="MLIVTGYTVYGRQVAHLLARTGPVAVLARPRPVYTGNNVY  
CAGYADPGLDRALNGVRAAFVLTNSATPPDDRFAPAAAREAGVRHLVKLSMLAVE  
PGADDFITRRQRENERAVRBSGLDWTELRARTEMSNTLSMAPALRSQGVVVALGTSP  
VACVDPDIEVAVARALTRPGHEBAVALGPEALITRAOQTARISEVLSRREEEG  
LEARBEOMLRKYPVPADAFLESERQRAKKAQVPTVAVETGRRPARPRIMADHA  
EAFRSQSG"  
complement(1970..2914)  
/gene="scbA"  
complement(1970..2914)  
/gene="scbA"  
/function="involved in gamma-butyrolactone synthesis"  
/codon\_start=1  
/transl\_table=1  
/product="ScbA protein"  
/protein\_id="CAA07627.1"  
/db\_xref="GI:3425860"  
/db\_xref="UniProt/TREMBL:O86851"  
/translation="MPRAVYLINSASANSISQALRPVPMALVHRTVQDAFPYSWIP  
KGGDRFSVTAVLPDHPFPAFVHGDRDPLLIATLLOAMLVTHAGYGFVGHFLM  
ATLDYTLCHLHLGVSQVAVLEVAASQLEKFGQVQGVDAVARGALATGTA  
TTTRFSPQVYVRMGDEATPTASVGPAPVAPARAGTRDDEVTLSSASQODTWRLRY  
DTSHPTEFORNDHVPGMLLEAAROAACLVTGAPVPVSGTRFVRYAFAFDSPCWI  
QATVPPGPAAGLTIVRYTGHQDGLVLTLLSGPAFSG"  
3032..3679  
/gene="scbR"  
3032..3679  
/gene="scbR"  
/codon\_start=1  
/transl\_table=1  
/product="gamma-butyrolactone binding protein"



```

/protein_id="CA007628.1"
/db_xref="GI:3425861"
/db_xref="GOA:086852"
/db_xref="UniProt/TREMBL:O86852"
/translation="MAKODRAIRRTQITLIDAAQVFEKQYQAAITTEILKAVGTVK
ALYFPOSKEELAGVPDQAEPPQVAPPEQRIINISSAATRRAPPSIGYAMTKGAVDTL
LSMOQAHGCDLRGPFRRMHTTLKLNQAENELLPHVYTTSSADLYVTPGSIQV
VSQTVSDYQDLERHYALQKAILPAIAPSVLAALDISEKGAELALAPLTKD"
gene
complement(3795..4529)
/gene="orfX"
complement(3795..4529)
/gene="orfX"
complement(3795..4529)
/gene="orfX"
/codon_start=1
/translation="3-ketoacyl-ACP/CoA reductase"
/product_id="CA007629.1"
/db_xref="GI:3425862"
/db_xref="GOA:086853"
/db_xref="UniProt/TREMBL:O86853"
/translation="MTGSSRGIGIGIALRLADALVAHYGSSAAARETETIRSS
GGQALAIKAEIGVVDAAALYAAPDAGWEGEVPEDILVNNAGVSGSGRTITVTE
VFDRLVAVNVAAPLVOHGLKRLDGRRIINISSAATRRAPPSIGYAMTKGAVDTL
TLAAROLGERGITVNAVAPGVETDMMARRRQTPPEAAALAAANSVNRIGRPDIAD
VVAFLADSDSRWITGVYDAGCTIL"
complement(4746..6446)
/gene="orfZ"
complement(4746..6446)
/gene="orfZ"
/codon_start=1
/translation="hypothetical protein"
/product_id="CA007630.1"
/db_xref="GI:3451475"
/db_xref="UniProt/TREMBL:O86854"
/translation="MMVVRRCGLVGRRRRRRKASRPBGAPAPAPWSSPPAARORA
GGPRTADARAVVDPAAGARLTPPEADQAEPPIDAGQATSAEGGATMAHPAPS
PGTAHABPASSPGADTQGSVAVRGTVSSAPPAKAPRSMAPDDGDRTFPLP
PPEAAAGPFAVATWGNAMVTLBEGALDARLRGCGYAEHGVDAITVPGVILVY
VRSGRSPYRQVRLRTAGDSMDRFLDAVERGHTALDGLPHSLADLADRGVY
LTPRGDLAPRCSPDSCGPKKHAALCYOPARILDDPFLILLRGSGEALLDALS
LRNARARAAQDRGPRPLPGVRAGALCTOPARILDDPFLILLRGSGEALLDALS
PYPPDLATDAARAHALITRRDPPGGLTLMQDAVRLAARPGSLTATGTAALTA
SLAAGRDTELARAAVAMRGGLAGLDVLEBMDPDPAGFRDAPPMMLAADIPAR
PMRRLTLTPRHVOLRLGRTGLMYAYSELGRBDMWPRGTPLDIPVGLTGLGCPDP
"
ORIGIN
Query Match 88.9%; Score 16; DB 1; Length 6727;
Best Local Similarity 88.9%; Pred. No. 3.7e+03;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Cy 1 GACCAAGTSCCGGCATG 18
Db 2224 GACCACGTACCGGCATG 2207

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
12000953
REFERENCE
AUTHORS
Bentley,S.D.
JOURNAL
COMMENT
FEATURES
SOURCE
1..290850
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
97..1101
/gene="SCO6257"
/note="synonym: SCAH10.22"
97..1101
/gene="SCO6257"
/note="SCO6257"
binding lipoprotein. len: 334 aa; similar to TR:CA841563
(EMBL:AL049727) Streptomyces coelicolor putative secreted
soluble binding protein. 337 aa; fasta scores: opt: 1003
z-score: 1130.9 E(0); 49.8% identity in 325 aa overlap,
D-ribiose-binding protein precursor RbsB. 305 aa; fasta
scores: opt: 206 z-score: 238.7 E(1); 6.3e-06; 23.0%
identity in 235 aa overlap and to TR:CA41563 (EMBL:SC9B1)
Streptomyces coelicolor SC9B1.17c. 337 aa; fasta scores:
opt: 1003 z-score: 1045.2 E(1); 0; 49.8% identity in 325 aa
overlap. Contains match to Pfam entry PF00532
Peripla_BP_1like. Periplasmic binding proteins and lacti
family and a match situated in the correct position to
Prosite entry PS00013 Prokaryotic membrane lipoprotein
lipid attachment site. Contains also possible N-terminal
region signal peptide sequence"
/codon_start=1
/translation="putative ABC transport system sugar binding
lipoprotein"
/product_id="CA860176.1"
/db_xref="GI:6273662"
/db_xref="UniProt/TREMBL:O9RKT5"
/translation="MAAPRTWVIGIALAGALSVLAGGSSNGKRAEDAPKASAEGR
AVDTPRTFAMITHSQDGTFFWDIVVSGAQAQVKNINLRYSHDAEQQAQVANA
IDKRVGDIITLAKPDMKSAALRAHAGIPIVTVNSGSEKSEFGALTHVGDETTA
GKAVGELNRRGQAVCVLHEQGNVGHQSCDGVETKPGQVRLVNLVTSMDVGS
AIAKIQTDKRSVAVVTLGAPVADTVKAKGAGSKAEIDTPILNAKVAAGLADGTIG
FANVDQDYLGAYAVDLMVLYKNADVLGGGRVPLVGPQIITDQDAALADLYTRGRTR
"
133..165
/gene="SCO6257"
/note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
241..966
/gene="SCO6257"
/note="Pfam match to entry PF00532 Peripla_BP_1like,
Periplasmic binding proteins and lacti family., score
13.80, E-value 2.1e-06"
1098..2138

```

/gene="SC06258"  
/note="synonym: SCAH10.23"  
1098. .2138  
/gene="SC06258"  
/note="SCAH10.23, probable ABC transport system sugar permease, len: 346 aa; similar to TR:O68120 (EMBL:AF010496) Rhodobacter capsulatus ribose transport system permease protein RbsC; fasta scores: opt: 445 z-score: 508.0 E(): 6.4e-21; 30.5% identity in 351 aa overlap and to SW:RBSB\_ECOLI1 (EMBL:L10328) Escherichia coli ribose transport system permease protein RbsC; len: 321 aa; fasta scores: opt: 286 z-score: 329.9 E(): 5.3e-11; 30.1% identity in 329 aa overlap. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/transl\_table=11  
/product="putative ABC transport system sugar permease"  
/protein\_id="CAB60177.1"  
/db\_xref="GI:6273663"  
/db\_xref="GOA:Q9RKT4"  
/db\_xref="UniProt/TREMBL:Q9RKT4"  
/translation="MNATQVKAADERILQTSRLRLGLRPELGSVGAIAVFEFAFADSLHASTLVLYASTIGIMAVPALIMIGFEPDLSAGVMTSSALVSMSYOMTANVAVGVVSLITLALGANGFMTLTKLPSPIITLGTMLTGMLGPTKLVDTGVSTKIDMEGPPSAQDVFASTITIGVGFVTLIMWLALVAVSMILTRAGNMFVAVGNKDAARAAGVPAVAKTIGLVNIGVGFAMISQHLFSYDVAVSGEVELELYIAAVVGCLITGCGYSAVSAVGAFIFQMTSKIVFAEMNDWFEFLGAMLLATLILNAAVRAEATK"  
2124. .2127  
/gene="SC06258"  
2135. .2926  
/gene="SC06259"  
/note="synonym: SCAH10.24"  
2135. .2926  
/gene="SC06259"  
/note="SCAH10.24, probable ABC sugar transport ATP binding protein, len: 263 aa; similar to TR:O68121 (EMBL:AF010496) Rhodobacter capsulatus ribose transport ATP-binding protein, 305 aa; fasta scores: opt: 612 z-score: 706.1 E(): 5.9e-32; 45.6% identity in 217 aa overlap, to SW:RBSA\_ECOLI1 (EMBL:M13169) Escherichia coli ribose transport ATP-binding protein RbsA, 501 aa; fasta scores: opt: 509 z-score: 585.6 E(): 3e-25; 31.5% identity in 254 aa overlap and to TR:CA15787 (EMBL:SC7B7) Streptomyces coelicolor ST787.07, 260 aa; fasta scores: opt: 592 z-score: 521.2 E(): 1e-23; 43.1% identity in 246 aa overlap. Contains Pfam match to entry PF00005 ABC tran. ABC transporter and Prosite PS00017 ATP/GTP-binding site motif A (P-loop) and PS00211 ABC transporters family signature"  
/codon\_start=1  
/transl\_table=11  
/product="probable ABC sugar transport ATP binding protein"  
/protein\_id="CAB60178.1"  
/db\_xref="GI:6273664"  
/db\_xref="GOA:Q9RKT3"  
/db\_xref="UniProt/TREMBL:Q9RKT3"  
/translation="MTRNEDRTALVELSGVKNYGVNVALBEGVSLFVAGRTCYLGDNAGKSTLIKTIAGLHQHGDGTLSDGETRSLSPREALDRGAIATVDDILPMLPVMNPFILGSRPRGVAPFKRMVDNHRKRTTHALMLRDLRVDDITLSSGGEQCV AIAAAVYFGAKVLVDEPTAALGVQSGVLYLVAAADQGLGVLLTHNPHAYLVGDRFVLKRGTVAAQNRDREVTLDLQMAAGSTELDDLRHLEERG"  
2240. .2815  
/gene="SC06259"  
/note="Pfam match to entry PF00005 ABC tran, ABC transporter, score 168.70, E-value 9.6e-47"  
2261. .2284  
/gene="SC06259"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
2388. .2632  
/gene="SC06259"  
/note="PS00211 ABC transporters family signature"

gene 3042. .4190  
/gene="SC06260"  
/note="synonym: SCAH10.25"  
3042. .4190  
/gene="SC06260"  
/note="SCAH10.25, possible sugar kinase, len: 382 aa; similar to SW:GLK\_STRCO (EMBL:X65932) Streptomyces coelicolor glucokinase (EC 2.7.1.2) (glucose kinase) Glk, 317 aa; fasta scores: opt: 1022 z-score: 1142.9 E(): 0; 46.6% identity in 311 aa overlap and to TR:CAB51974 EMBL:SC6E10 Streptomyces coelicolor SC6E10.20C, 317 aa; fasta scores: opt: 1022 z-score: 1035.4 E(): 0; 46.6% identity in 311 aa overlap. Contains a match to Pfam entry PF00480 ROK, ROK family and Prosite entry PS01125 ROK family signature"  
/codon\_start=1  
/transl\_table=11  
/product="putative sugar kinase"  
/protein\_id="CAB60179.1"  
/db\_xref="GI:6273665"  
/db\_xref="GOA:Q9RKT2"  
/db\_xref="UniProt/TREMBL:Q9RKT2"  
/translation="MSTYRDFTAIGSRRAVLRVTVGTRRRSHLTAPVPVIGIDIGTKVMAGVADAGNGLIEKLRTEPPDKSPKVEDTIVELDLSDHDAVAGIGAA GTMDADNRVLPAHLISWRNEPLRDRTAGRLAVPVLYVNDANLAAEMRFGGRGD HLWITLITGIGAILIEDQYKRGKRTVNAEFGMVVPPGHRCPGCRRCWYSSG NALVREARELAAADS PVAAYGIIEHVKSISDITGPMITELARBDACVELDQIGM LGVAGIANIAAALDPSCFVIGGVSAADLLIGPARDFKQLTGRGPARIVRAOL GPEAGVGAADLSRLVARRPRARRRVERERYERYEAARESEBSL"  
3159. .3704  
/gene="SC06260"  
/note="Pfam match to entry PF00480 ROK, ROK family, score 215.50, E-value 5.4e-63"  
3540. .3623  
/gene="SC06260"  
/note="PS01125 ROK family signature"  
4244. .4247  
4259. .4768  
/gene="SC06261"  
  
misc\_feature  
RBS  
gene  
  
Query Match 88.9%; Score 16; DB 1; Length 290850;  
Best Local Similarity 88.9%; Pred. No. 1.8e+03;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 GACCACGTSCCGGCGATG 18  
||||| :|||  
Db 12783 GACCACGTACCGGCGATG 12800  
  
RESULT 9  
BD139044  
LOCUS 513 bp DNA linear PAT 18-SEP-2002  
DEFINITION Virulence-associated nucleic acid sequences and uses thereof.  
ACCESSION BD139044  
VERSION BD139044.1 GI:23233989  
KEYWORDS JP 2002505849-A/92.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 513)  
Ausbel,F., Goodman,H.M., Rahme,J.G., Miklos,S.M., Tan,M.W., Cao,H., Drenkard,E. and Tsongalis,J.  
Virulence-associated nucleic acid sequences and uses thereof  
Patent: JP 2002505849-A 92 26-FEB-2002;  
THE GENERAL HOSPITAL CORP  
OS Pseudomonas aeruginosa  
PN JP 2002505849-A/92  
PD 26-FEB-2002  
PF 25-NOV-1998 JP 2000522270  
PR 25-NOV-1997 US 60/066517  
PI FREDERICK AUSBEL, HOWARD M GOODMAN, LAURENCE G RAHME PI  
, SHALINA MAHAJAN MIKLOS,



ORGANISM Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 762)  
AUTHORS Aubel, F., Goodard, H.M., Rahme, L.G., Mahajan-Miklos, S., Tan, M.-W., Cao, H., Drenkard, E. and Tsongalis, J.  
TITLE Virulence-associated nucleic acid sequences and uses thereof  
JOURNAL Patent: US 6355411-A 153 12-MAR-2002;  
FEATURES Location/Qualifiers  
source 1..762  
/organism="Unknown"  
/mol\_type="unassigned DNA"  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 762;  
Best Local Similarity 83.3%; Pred. No. 8.6e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTCCGCGCATG 18  
|||||:|:|:|:|:|:|  
Db 303 GACCACGTCCGCGCATG 286  
RESULT 14  
AY060758 859 bp mRNA linear INV 08-NOV-2001  
LOCUS Drosophila melanogaster GH20904 full length cDNA.  
ACCESSION AY060758  
VERSION AY060758.1 GI:16768173  
KEYWORDS FLI CDNA.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 859)  
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Aghayani, A., Carlson, J., Champagne, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Li, P., Liao, G., Miranda, A., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M. and Ceiniker, S.  
TITLE Direct Submission  
JOURNAL Submitted (30-OCT-2001) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
COMMENT Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720  
This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).  
FEATURES Location/Qualifiers  
source 1..859  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/map="50C9-50C11"  
1..859  
/gene="CG6305"  
/note="alignment with genomic scaffold AB003817"

CDS  
/db\_xref="FLYBASE:FBgn0033869"  
208..744  
/gene="CG6305"  
/note="Longest ORF"  
/codon\_start=1  
/product="GH20904p"  
/protein\_id="AA128306.1"  
/db\_xref="GI:16768174"  
/db\_xref="FLYBASE:FBgn0033869"  
/translation="MMSFKSFLSLSLAISAYAYAQAPGSNAYLPTKNGYDSE  
PKTPFKPGRPAAPGPPAPGPPGPDHVPQMPYDPFEAYVDDPERA  
NDYAHKASDGDVVTGEYRVQMPDGRQIVRYADWKTGYHADVSVEGATYVQGPQP  
GARGGGGGAGAGAGGYKY"  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 3; Length 859;  
Best Local Similarity 83.3%; Pred. No. 8.4e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTCCGCGCATG 18  
|||||:|:|:|:|:|:|  
Db 454 GTCACGTGCCCGGCATG 471  
RESULT 15  
AR480461/c 1008 bp DNA linear PAT 14-MAY-2004  
LOCUS AR480461  
DEFINITION Sequence 1365 from patent US 6699703.  
ACCESSION AR480461  
VERSION AR480461.1 GI:47239423  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 1008)  
AUTHORS Doucette-Stamm, L., Bush, D., Zeng, Q., Oppertman, T. and Houseweart, C.E.  
TITLE Nucleic acid and amino acid sequences relating to Streptococcus pneumoniae for diagnostics and therapeutics  
JOURNAL Patent: US 6699703-A 1365 02-MAR-2004;  
FEATURES Location/Qualifiers  
source 1..1008  
/organism="Unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 1008;  
Best Local Similarity 83.3%; Pred. No. 8.2e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTCCGCGCATG 18  
|||||:|:|:|:|:|:|  
Db 556 GACCACGTGCCCGGCATG 539  
RESULT 16  
AX567274/c 1188 bp DNA linear PAT 29-NOV-2002  
LOCUS AX567274  
DEFINITION Sequence 473 from Patent WO02077021.  
ACCESSION AX567274  
VERSION AX567274.1 GI:26001699  
KEYWORDS  
SOURCE Streptococcus pneumoniae  
Streptococcus pneumoniae  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1  
AUTHORS Maignani, V., Tettelin, H. and Fraser, C.  
TITLE Streptococcus pneumoniae proteins and nucleic acids  
JOURNAL Patent: WO 02077021-A 473 03-OCT-2002;  
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)  
FEATURES Location/Qualifiers  
source 1..1188

/organism="Streptococcus pneumoniae"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1313"

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 1188;  
Best Local Similarity 83.3%; Pred. No. 7.9e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 739 GACCACGTGCCGCATG 722

RESULT 17

E30250 1408 bp DNA linear PAT 18-JUN-2001  
LOCUS Process for producing biotin.  
DEFINITION E30250  
ACCESSION E30250.1 GI:13021423  
VERSION JP 199276175-A/1.  
KEYWORDS Sphingomonas sp.  
SOURCE Sphingomonas sp.  
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;  
Sphingomonadaceae; Sphingomonas.

REFERENCE 1 (bases 1 to 1408)  
AUTHORS Tomoyasu, K., Fujio, S. and Masatoshi, S.  
TITLE Process for producing biotin  
JOURN. Patent: JP 199276175-A 1 12-OCT-1999;  
SUMITOMO CHEM CO LTD  
OS Sphingomonas sp.  
PN JP 199276175-A/1  
PD 12-OCT-1999  
PF 31-MAR-1998 JP 1998086972  
PR

PI TOMOYASU KAMABE, FUJIO SUZUMOTO, MASATOSHI SIMIZU PC  
C12N15/09, C12N1/21, C12P17/18// (C12N15/09, C12R1:01), (C12P17/18, PC  
C12R1:19),  
PC C12N15/00, (C12N15/00, C12R1:01)  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT CDS Location/Qualifiers  
1..1408  
/organism="Sphingomonas sp."  
/mol\_type="genomic DNA"  
/db\_xref="taxon:28214"

FEATURES  
source

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 1408;  
Best Local Similarity 83.3%; Pred. No. 7.7e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1057 GACCTCGTCCGCGCATG 1074

RESULT 18

E30254 1408 bp DNA linear PAT 18-JUN-2001  
LOCUS Process for producing biotin.  
DEFINITION E30254  
ACCESSION E30254.1 GI:13021427  
VERSION JP 199276175-A/5.  
KEYWORDS Sphingomonas sp.  
SOURCE Sphingomonas sp.  
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;  
Sphingomonadaceae; Sphingomonas.

REFERENCE 1 (bases 1 to 1408)  
AUTHORS Tomoyasu, K., Fujio, S. and Masatoshi, S.  
TITLE Process for producing biotin

JOURNAL Patent: JP 1999276175-A 5 12-OCT-1999;  
SUMITOMO CHEM CO LTD  
COMMENT OS Sphingomonas sp.  
PN JP 199276175-A/5  
PD 12-OCT-1999  
PF 31-MAR-1998 JP 1998086972  
PR

PI TOMOYASU KAMABE, FUJIO SUZUMOTO, MASATOSHI SIMIZU PC  
C12N15/09, C12N1/21, C12P17/18// (C12N15/09, C12R1:01), (C12P17/18, PC  
C12R1:19),  
PC C12N15/00, (C12N15/00, C12R1:01)  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT CDS Location/Qualifiers  
1..1408  
/organism="Sphingomonas sp."  
/mol\_type="genomic DNA"  
/db\_xref="taxon:28214"

FEATURES  
source

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 1408;  
Best Local Similarity 83.3%; Pred. No. 7.7e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1057 GACCTCGTCCGCGCATG 1074

RESULT 19

AR215256 1408 bp DNA linear PAT 25-SEP-2002  
LOCUS Sequence 5 from patent US 6410293.  
DEFINITION AR215256  
ACCESSION AR215256.1 GI:23313427  
VERSION AR215256.1 GI:23313427  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1408)  
AUTHORS Mukumoto, F., Nishio, S., Akimaru, J. and Mitsuada, S.  
TITLE DNA fragments containing biotin biosynthease gene and use of the  
same  
JOURNAL Patent: US 6410293-A 5 25-JUN-2002;  
FEATRES Location/Qualifiers  
1..1408  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 1408;  
Best Local Similarity 83.3%; Pred. No. 7.7e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1057 GACCTCGTCCGCGCATG 1074

RESULT 20

AR215257 1408 bp DNA linear PAT 25-SEP-2002  
LOCUS Sequence 7 from patent US 6410293.  
DEFINITION AR215257  
ACCESSION AR215257.1 GI:23313428  
VERSION AR215257.1 GI:23313428  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1408)  
AUTHORS Mukumoto, F., Nishio, S., Akimaru, J. and Mitsuada, S.

TITLE DNA fragments containing biotin biosynthesase gene and use of the same

JOURNAL Patent: US 6410293-A 7 25-JUN-2002;

FEATURES Location/Qualifiers

source 1..1408

/organism="unknown"

/mol\_type="genomic DNA"

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 1408;

Best Local Similarity 83.3%; Pred. No. 7.7e+03;

Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGGCATG 18

Db 1057 GACCTCGTCCGCGGCATG 1074

RESULT 21

LOCUS NCA250111 1794 bp DNA linear PLN 14-FEB-2003

DEFINITION Nitella capillaris partial 18S rRNA gene.

ACCESSION AJ250111

VERSION AJ250111.1 GI:6688855

KEYWORDS 18S ribosomal RNA; 18S rRNA gene.

SOURCE Nitella capillaris

ORGANISM Nitella capillaris

Eukaryota; Viridiplantae; Streptophyta; Charales; Characeae; Nitella.

REFERENCE 1

AUTHORS Marin, B. and Melkonian, M.

TITLE Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequence comparisons

JOURNAL Protist 150 (4), 399-417 (1999)

MEDLINE 20177256

PUBMED 10714774

REFERENCE 2 (bases 1 to 1794)

AUTHORS Marin, B.

TITLE Direct Submission

JOURNAL Submitted (01-OCT-1999) Martin B., Botanisches Institut, Lehrstuhl I, Universitaet zu Koeln, Gyrhofstrasse 15, 50931 Koeln, GERMANY

FEATURES source

1..1794

/organism="Nitella capillaris"

/mol\_type="genomic DNA"

/db\_xref="taxon:105230"

/country="Germany"

1..1794

/gene="18S rRNA"

<1..1794

/gene="18S rRNA"

/product="18S ribosomal RNA"

ORIGIN

Query Match 86.7%; Score 15.6; DB 8; Length 1794;

Best Local Similarity 83.3%; Pred. No. 7.3e+03;

Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGGCATG 18

Db 696 GACCACTGTCGCGGCATG 679

RESULT 22

LOCUS AX656460 2000 bp DNA linear PAT 22-MAR-2003

DEFINITION Sequence 6330 from Patent WO03000898.

ACCESSION AX656460

VERSION AX656460.1 GI:29159274

KEYWORDS

SOURCE Oryza sativa

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1

AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M., Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.

TITLE Plant genes involved in defense against pathogens

JOURNAL Patent: WO 03000898-A 6330 03-JAN-2003;

SYNGENTA Participations AG (CH)

FEATURES Location/Qualifiers

source 1..2000

/organism="Oryza sativa"

/mol\_type="unassigned DNA"

/db\_xref="taxon:4530"

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 2000;

Best Local Similarity 83.3%; Pred. No. 7.2e+03;

Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGGCATG 18

Db 1089 GGCACAGTCCGCGGCATG 1106

RESULT 23

LOCUS BD139028 2048 bp DNA linear PAT 18-SEP-2002

DEFINITION Virulence-associated nucleic acid sequences and uses thereof.

ACCESSION BD139028

VERSION BD139028.1 GI:23233973

KEYWORDS JP 2002505849-A/76.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 2048)

AUTHORS Ausubel, F., Goodman, H.M., Ratme, L.G., Miklos, S.M., Tan, M.W., Cao, H., Drenkard, E. and Tsongalis, J.

TITLE Virulence-associated nucleic acid sequences and uses thereof

JOURNAL Patent: JP 2002505849-A 76 26-FEB-2002;

THE GENERAL HOSPITAL CORP

COMMENT OS Pseudomonas aeruginosa

PN JP 2002505849-A/76

PD 26-FEB-2002

PR 25-NOV-1998 JP 2000522270

PI 60/066517

PI FREDERICK AUSUBEL, HOWARD M GOODMAN, LAURENCE G RAHME PI

SHALINA MAHAJAN MIKLOS, ELIANA DRENKARD, JOHN TSONGALIS PC

PI MAN WAH TAN, HUI CAO, ELIANA DRENKARD, JOHN TSONGALIS PC

GI2N15/09, A61K45/00, A61P31/04, C07K14/21, C12Q1/02, C12Q1/68, PC

G01N33/15,

PC G01N33/50//C12P21/02, (C12N15/09, C12R1.385), C12N15/00, (C12N15/00, C12R1.385)

CC Virulence-associated nucleic acid sequences and uses thereof

CC Key

FT source 1..2048

FT Location/Qualifiers

FT

FEATURES source

1..2048

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/db\_xref="taxon:287"

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 2048;

Best Local Similarity 83.3%; Pred. No. 7.1e+03;

Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGGCATG 18

Db 1069 GACCACTGTCGCGGCATG 1086

RESULT 24  
LOCUS AR198851 2048 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 136 from patent US 6355411.  
ACCESSION AR198851  
VERSION AR198851.1 GI:20248925  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
FEATURES  
REFERENCE 1 (bases 1 to 2048)  
AUTHORS Ansubel, P., Goodman, H.M., Rahme, L.G., Mahajan-Miklos, S., Tan, M.-W., Cao, H., Drenkard, E. and Tsongalis, J.  
TITLE Virulence-associated nucleic acid sequences and uses thereof  
JOURNAL Patent: US 6355411-A 136 12-MAR-2002;  
FEATURES  
source  
1. 2048  
Location/Qualifiers  
/organism="unknown"  
/mol\_type="unassigned DNA"  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 2048;  
Best Local Similarity 83.3%; Pred. No. 7.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTSCCGGCATG 18  
|||||:|:|:|  
Db 1069 GACCACGTSCCGGCATG 1086  
RESULT 25  
LOCUS CQ789114 2395 bp DNA linear PAT 29-MAR-2004  
DEFINITION Sequence 205 from Patent EP1400592.  
ACCESSION CQ789114  
VERSION CQ789114.1 GI:45822682  
KEYWORDS  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1  
AUTHORS Kunesh, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Barash, S.C., Fannon, M. and Dougherty, B.A.  
TITLE Streptococcus pneumoniae polynucleotides and sequences  
JOURNAL Patent: EP 1400592-A 205 24-MAR-2004;  
FEATURES  
source  
1. 2395  
Location/Qualifiers  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 2395;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTSCCGGCATG 18  
|||||:|:|:|  
Db 1257 GACCACGTGCCGACATG 1240  
RESULT 26  
LOCUS AR218973 2395 bp DNA linear PAT 25-SEP-2002  
DEFINITION Sequence 205 from patent US 6420135.  
ACCESSION AR218973  
VERSION AR218973.1 GI:23319907  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2395)

AUTHORS Kunesh, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C., Fannon, M.R. and Dougherty, B.A.  
TITLE Streptococcus pneumoniae polynucleotides and sequences  
JOURNAL Patent: US 6420135-A 205 16-JUL-2002;  
FEATURES  
source  
1. 2395  
Location/Qualifiers  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 2395;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTSCCGGCATG 18  
|||||:|:|:|  
Db 1257 GACCACGTGCCGACATG 1240  
RESULT 27  
LOCUS BD003885 2395 bp DNA linear PAT 31-JAN-2002  
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.  
ACCESSION BD003885  
VERSION BD003885.1 GI:18631846  
KEYWORDS JP 2001501833-A/205.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 2395)  
AUTHORS Kunesh, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Barash, S.C., Fannon, M. and Dougherty, B.A.  
TITLE Polynucleotide of Streptococcus pneumoniae and sequence  
JOURNAL Patent: JP 2001501833-A 205 13-FEB-2001;  
COMMENT HUMAN GENOME SCIENCES INC  
OS Unidentified  
PN JP 2001501833-A/205  
PD 13-FEB-2001  
PF 30-OCT-1997 JP 1998520718  
PR 31-OCT-1996 US 60/029960  
PI CHARLES A KUNESH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI STEVEN C BARASH,  
MICHAEL FANNON, BRIAN A DOUGHERTY  
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19, C12N1/21, PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC G06F15/40  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key  
FT source  
1. 2395  
Location/Qualifiers  
/organism="unidentified".  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 2395;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTSCCGGCATG 18  
|||||:|:|:|  
Db 1257 GACCACGTGCCGACATG 1240  
RESULT 28  
LOCUS HSH12B 3233 bp DNA linear PRI 20-MAY-1996  
DEFINITION Homo sapiens H12B gene.  
ACCESSION X81326

VERSION	X61326.1 GI:587211		
KEYWORDS	repeated element.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Rybkov,A.P., Prosnjak,M.I., Kupriyanova,N.S., Netchvolodov,K.K. and Limborska,S.A.		
TITLE	A new multi-locus DNA fingerprinting probe: K25		
JOURNAL	DNA Res. 2 (3), 151-152 (1995)		
MEDLINE	96038269		
PUBMED	8581743		
REFERENCE	2 (bases 1 to 3233)		
AUTHORS	Prosnjak,M.I.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-SEP-1994) M.I. Prosnjak, Inst of Molecular Genetics, Russian Academy of Sciences, Kurchatov sq, 123182 Moscow, RUSSIA		
FEATURES	Location/Qualifiers		
source	1..3233		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/clone="H12B"		
	/haplotype="diploid"		
	/cell_type="lymphocytes"		
	/clone_id="human genomic DNA in lambda EMBL-3 phage"		
	/dev_stage="adult"		
repeat_region	38..>540		
	/evidence=experimental		
repeat_region	576..>840		
	/evidence=experimental		
ORIGIN			
Query Match	86.7%; Score 15.6; DB 9; Length 3233;		
Best Local Similarity	83.3%; Pred. No. 6.6e+03;		
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;			
Qy	1 GACCACTGTCCCGGCATG 18		
	:  :		
Db	829 GACCACTGTCCCGGCATG 846		
RESULT 29			
AF227905	4927 bp mRNA linear PRI 30-APR-2001		
LOCUS	Homo sapiens UDP-glucose:glycoprotein glucosyltransferase 1		
DEFINITION	precursor, mRNA, complete cds.		
ACCESSION	AF227905		
VERSION	AF227905.1 GI:7670745		
KEYWORDS			
SOURCE	.		
ORGANISM	Homo sapiens (human)		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 4927)		
REFERENCE	Arnold,S.M., Fessler,L.I., Fessler,J.H. and Kaufman,R.J.		
AUTHORS	Two homologues encoding human UDP-glucose:glycoprotein glucosyltransferase differ in mRNA expression and enzymatic activity		
TITLE	Biochemistry 39 (9), 2149-2163 (2000)		
JOURNAL			
MEDLINE	20160491		
PUBMED	10694380		
REFERENCE	2 (bases 1 to 4927)		
AUTHORS	Arnold,S.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-JAN-2000) Biological Chemistry and Howard Hughes Medical Institute, University of Michigan, 1150 West Medical Center Drive, Ann Arbor, MI 48109-0650, USA		
FEATURES	Location/Qualifiers		
source	1..4927		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		

```

CDS
/db_xref="tacon:9606"
/tissue_type="liver"
/dev_stage="fetal"
39..4706
/note="HUGT1"
/product="UDP-glucose:glycoprotein glucosyltransferase 1
precursor"
/protein_id="AA66232.1"
/db_xref="GI:7670746"
/translation="MGCCGASGACAAGALPTVGCMKVLVVLTVLMFSSVKDSS
KATTTSTTKWFTPTLLLEASERFLAESDSOKFNNFVASQNGISDHDGDYSYHAI
LEBAFQSLSPICQNLKFKCISLRSGATQAPQIQAADDEPPGCSNFFVHGKTEE
SDTLLEALFLASERPKPLFKGDRFPSSNPEBPVIFYEISSEFSNHRQLISKS
NACKINVFRRYVIFNPERKEPVLSGVGEVLAISTEKADDTQVKTQTEVNTYIGEN
DPIDVEQGLFKGLRDLPDLEQQLKELRHGLVESTNEMAPLKWQGLDSFQTAARL
LASPEVALVFMKDLSONPFTKARALITKIVASSLRFEVENQGFYKGTGLPGDSA
LFTINLHMDLDTQDIFSLPVLINREAVMEGLRHGLLEGSLNHVLLKNTOPSADAY
VDIRSPALSWNNLLEVDNRYSNWSPSICELLRPPEPVITQIRKLNHWIYIDPAHEH
TVBLMNTAMFELSNHLPRIIGTIFVNDSEDVDGMQDAVVALRATNYAQEDVDYH
AFQTLITYNKVRGKEVKVEHVAVSLEKKPYEVANSILGIDSAVRNKRKAGYIE
QTGGLPLPVLFNGMPEREQLDPDELETTIMKILETTFPQAAVVLGLPHQDQYH
EYKINQNVVRINSRILFLTAERDVLDTASNNEFVDYARFTLDSGKTAAVANSMM
YLTKKMSKEIYDSDIRIPVTMIVGDPSGROLVDAIKHKSNNVRIINNKN
PAKEISTENQISRAITMAALQOTQSNAAKMPITMAKEGAALAAAGADIAEPGVM
DPSLFEKVEFSSKQDFTLSHAAYTCRDVLAKKGQRAVISGRITIGLEBELFNQDDP
HLLENITLKTSQOKIKSHIQOLRVEEDVADLVKQVALLSAQKQPRLEYOFEEDR
HSAIKLPRKEGETYFDVAVVDPYTRBAQSLPLVLAQILNNMLVFVNNCSKLSL
MPLKSPFAYVLEPIEPTSDNSFAKGP IAKELDMPQSPPLFTLINTPESWMBEVRTP
YDLDNITLFEVDVVAAYEVELEVLLLEGHGCDITTGQPPGLOFTITGSAWVVDITIT
VMANTYGFQKADNPGAMITLRLKRSSEDIYRIYSHDQDSPPDADEVYIIVNNKSKIT
IKAVQKKADMVNEDLSDGTSSESGFMQSFKMGFTGQTEEVKQDKDQILINFSVA
SGHLYERPLRIMMLSVLKNKTIPVKEPFLNKYISPTKEFI PYMANEYNQYELVQYRK
WPRMLHQOTEKORILMGYKTLFLDVLPLVLVDKFLFVDAQIVTDLKEKLEDFMLDAQ
PYGTTFPCDSRREMDGYRFWMSGYMAHLAGRKYHIALYVVDLKEKPKIAAGGRLEK
OYQGLSDPPNSLMDQLDLPNNMHQVPIKSLPQEMLCETWCDASKEKAKTIDLCLNCL
NPMTKERKLEAARIVIEEMQDYDEITIQQLIRPQKEKETALYKEKTKPSREBPQKRR
BEL"
39..164
165..4703
/product="UDP-glucose:glycoprotein glucosyltransferase 1"

Query Match      86.7%; Score 15.6; DB 9; Length 4927;
Best Local Similarity 83.3%; Pred. No. 6.1e+03;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy      1 GACCAAGTSCGCGCATG 18
      |||||.:|||
Db      24 GACCACGCGCCGCGCATG 41

RESULT 30
AX405692 LOCUS      5042 bp      DNA      linear      PAT 14-JUN-2002
DEFINITION Sequence 107 from Patent WO0222660.
ACCESSION AX405692
VERSION AX405692.1 GI:21438823
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Tang,Y.T., Liu,C., Zhou,P., Aouni,V., Zhang,J., Zhao,Q.A., Ren,F.,
TITLE Xue,A.J., Yang,Y., Wehman,T. and Demanac,R.T.
JOURNAL Novel nucleic acids and polypeptides
PUBLISHED Patent: WO 0222660-A 107 21-MAR-2002;
FEATURES
SOURCE location/Qualifiers
1..5042
/organism="Homo sapiens"
/mol_type="unassigned DNA"

```



## CDS

```
/db_xref="taxon:9606"  
164..4831  
/note="unnamed protein product"  
/codon_start=1  
/protein_id="CAD34799.1"  
/db_xref="GI:21438824"  
/translation="MGCKGDASGACAAGALPTGVCYKMGVLVLTVMLFSSVKADS  
KAITTSITTKMFTPLLEASERPLAEDSQEFNPFVSAQNISSDHGTDYSYHAI  
LEAFOPLSPLOQNLPKFCLSRSYSAIOAFQOIADEPPEGCNSFFSVHGKTCE  
SDTEBALLTASERPKPLPKGDHRYSSNPSPVIFYSEIGSEEFNFRQLISKS  
NAGKINVPFRHYIENPKPEPVYLSGVGVELAIKSTYKADDPQVGTETVNTVIGEN  
DPIDVEOGFLPKIRDLHPDLBGLKELRKLHVKSTENMAPLKYMLQDLSPQTAHI  
LASPEVALVVMKDLSONPPTKAAITKTAVSSELRTVEENQYFEKTYLGLQPADSA  
LFTNGLMDDLTDIDFSLFDVLRNEARVMGLHRLIGLSLHVLYKLNTQSEADYA  
VDLRSPAISVNNLEVDSTRYNSWSSLOELLPTFPGVIRQIRNLNNVFIYDPAE  
TTAELMTAMFPLSNHIPLRIGFIIVNDESDVGMQDAGVAIRAYNVAOEVDPH  
AFQTLPIYVNVKVTGEEKVKEHVSVLEKQPYVEVNSITIGIDSAVDNRKREAGYYE  
QTGGPLPVVLFNGMPFERQLDDELTTMKILRTTTFQRAVYLGELPHQDAV  
BYIMNQNVNVRINSRILTAARDYLDLTASNFEVDYARFTILDSQKTAAVANSNN  
YLTKGMSSEIYDDSFIRPTEWIVGDPSPSGRQLLYDAIKHKSNNVRISMNN  
PAKEISYENTOISRAIWAALQOTSNAAKNFTMAKEGAABALAGADIAEFSVGM  
DPSLPEKVEFSSKMDFILSHAVCRDVLYLKGORAVISNGRIIGPLEDSELFNODP  
HLENTILKTSQKIKSHIQQLRVEEDVASDVMKVYALLSAQKGPRIEYQFERE  
HSAIKLRKEGETFDVVAVVDVYREAQRLAPLLVLAQLINNRLRFNACSKLSD  
MPLKSPRYVLEPEISFTSDNSPAKGPIDAKFLDMPQSLPTLNLNTPESMWVSVRTP  
YLDNITYLEEVDSVAAEYLEYLLLEGHCYDITTCQPPRGLQFTLGTSANPVIYDTI  
VMANIGYFOLKANGAWILRLRKGSEDIYRISHDGDSPPDADEVIVIANPFSKI  
IKYVOKKADMVNEDLLDGTSENESEGFMSFKGFTGKTEEVKDDKDDIINISVA  
SGHLIERFLIMLISUKTKTTPYKFWFLKNTYLSPTPKEFIPLYMANEYNFOYELVOYK  
WPRMLHQTEKQRIWGYKILFLDVLPVVDKFLFVDAQIVRTDLKEILDVFLDGA  
PYGVTPECDSRRENDGYRFWKSQYMAHLAGRKYHISALYVDLKKFRKIAAGDRLG  
QYQGLSQDPSNLSLDDQLPNNMTHQVPIKSLPEWMLMCEMTCDDASKRAKTIIDLGN  
NPMTKEPKLEAAVRIPEWQDYDDEIKQLIRFOKEKETGALYKTKEPSREGPQKR  
EEL"
```

## ORIGIN

```
Query Match      86.7%; Score 15.6; DB 6; Length 5042;  
Best Local Similarity 83.3%; Pred. No. 6e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 GACCACTSCCGGCATG 18  
        |||||:|||||  
Db      149 GACCACGCGCCGCGCATG 166
```

```
Search completed: July 20, 2005, 16:45:08  
Job time : 903.429 secs
```

**THIS PAGE IS BLANK**